UiO : University of Oslo

Çağrı Erdem

Controlling or Being Controlled? Exploring Embodiment, Agency and Artificial Intelligence in Interactive Music Performance

Thesis submitted for the degree of Philosophiae Doctor

Department of Musicology
Faculty of Humanities

RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion

2022

Thesis submitted 2 November 2021
Thesis defended 10 May 2022

Advisors:

Professor Alexander Refsum Jensenius, University of Oslo
Professor Rolf Inge Godøy, University of Oslo
Professor Kyrre Glette, University of Oslo

Committee:

Professor Palle Dahlstedt, University of Gothenburg
Professor Gascia Ouzounian, University of Oxford
Associate Professor Jonna Vuoskoski, University of Oslo

Faculty of Humanities, University of Oslo

Cover: Hanne Baadsgaard Utigard.
Print production: 07-Media Oslo.

My CPU is a neural-net processor; a learning computer.
The more contact I have with humans, the more I learn.
– The Terminator

Abstract

How can we make music with artificial intelligence (AI) in the future? Unlike most studies on AI and music, this dissertation focuses on physical interaction and the ways in which the computer can respond to body movement. Based on experimental music practices, it argues that diversifying artistic repertoires in music-making is crucial for the future of music. Emphasis has been placed on realizing creative works and their evaluations in ecological environments. The exploration starts from an extensive literature review that sketches a broad picture of alternative control paradigms in the performing arts, different types of musical AI, and embodied approaches to human cognition. Then follows a methodological presentation and discussion structured around the four projects that the dissertation is focused on. The shared music–dance piece *Vrengt* demonstrates the musical possibilities of sonic microinteraction and provides a conceptual model of co-performance. The muscle-based instrument *RAW* implements various AI techniques to explore a chaotic instrumental behavior and automated interaction with an improvisation ensemble. A novel empirical study sheds light on how guitar players transform biomechanical energy into sound. The collected multimodal dataset is used as part of a modeling framework for "air performance." The coadaptive audiovisual instrument *CAVI* uses generative modeling to automate live sound processing and investigates expert improvisers' varying sense of agency. All in all, this dissertation stresses the importance of embodied perspectives when developing musical AI systems. It emphasizes an entwined artistic–scientific research model for interdisciplinary studies on performing arts, AI, and embodied music cognition.

Sammendrag

Hvordan vil vi lage musikk med kunstig intelligens (KI) i fremtiden? I motsetning til de fleste studier innen KI og musikk, fokuserer denne avhandlingen på fysisk interaksjon og på hvilken måte datamaskiner kan svare på kroppsbevegelser. Med utgangspunkt i en eksperimentell musikkpraksis, argumenteres det for at en utvidelse av det kunstneriske repertoaret er avgjørende for fremtidens musikk. Avhandlingen har vektlagt å realisere kreativt arbeide og evaluering i økologiske omgivelser. Utforskningen springer ut fra en omfattende litteraturgjennomgang som tegner opp et bredt bilde av alternative kontrollparadigmer i scenekunsten, ulike typer musikalsk KI og kroppslige tilnærminger til menneskelig kognisjon. Deretter følger en metodologisk presentasjon og diskusjon strukturert rundt de fire prosjektene avhandlingen springer ut i fra. Det delte musikk-dans-stykket *Vrengt* demonstrerer de musikalske mulighetene med lydlig mikrointeraksjon og tilbyr en modell for sam-spilling. Det muskelbaserte instrumentet *RAW* implementerer variasjoner av KI-teknikker for å utforske en kaotisk instrumentell oppførsel og automatisert interaksjon med et improvisasjonsensemble. En empirisk studie undersøker hvordan gitarister omgjør biomekanisk energi til lyd. Det innsamlete multimodale datasettet utgjør deler av et rammeverk for "luftspilling." Det koadaptive audiovisuelle instrumentet *CAVI* benytter generativ modellering for å automatisere lydprosessering i sanntid og undersøker profesjonelle improvisatørers varierende opplevelse av agency. Alt i alt, vektlegger denne avhandlingen viktigheten av kroppslige perspektiver for utviklingen av musikalske KI-systemer. Den fokuserer på en sammenvevd kunstnerisk-vitenskapelig forskningsmodell for interdisiplinære studier av scenekunst, KI, og kroppslig musikkognisjon.

Acknowledgements

Doing a Ph.D., which resulted in this dissertation, has been such a journey that had both grueling and euphoric extremes. “It’s like riding a roller coaster,” once said a wise friend, and he was right. I am particularly grateful to my main supervisor, who has guided me with rigor and enthusiasm during this long ride. This work would not be possible without you, Alexander. Every time I fell, you showed me some way out and helped me put myself together. I have learned a lot from you—and I am still learning. You are a true inspiration!

I am also grateful to my co-supervisors. We have had countless thought-provoking discussions with Rolf Inge on sound and philosophy, and the concepts and theories he has been working on for decades provided my project with solid departure points. With Kyrre, it is always a pleasure to ponder exciting engineering problems and to have geeky discussions on algorithms or embedded platforms. Thank you both for your friendliness, insights, and welcoming of all my confused and fuzzy states.

I also want to use this opportunity to go back in time and thank the people who have left a mark on me, helping me transform into who I am. As a music student, I have had some great teachers and mentors. Şevket Akıncı, Selen Gülün, Şenol Küçükyıldırım, Dave Tronzo, and Mick Goodrick were among the first ones who showed me the “other” ways of musicking. The music technological practice I am still working on today gained momentum afterward within a small community of artists, creative coders, and DIY enthusiasts in İstanbul. I began exploring many of the concepts and methods subject to this dissertation back in those times during days-long hackathons, jam sessions, events, and conversations. I am grateful to many I have interacted with in varying degrees of familiarity and intimacy, particularly to Görkem, Yurdal, İpek, Giray, all the A.I.D, Bahçe and Galata crews, and Utku & Anja from Multiversal. Here I must mention Erdem, who gave me my first Myo armband as a gift, and Musa, who accompanied me to several international events and conferences.

During my Master’s studies, I was fortunate to work with Anıl Çamcı, who not only immensely helped and motivated me for further graduate studies but also introduced me to the New Interfaces for Musical Expression (NIME) community. I remember the first time I browsed the NIME website and spent a great deal of time exploring the conference proceedings—this was when I first found out about Alexander’s work, which struck me immediately. I did not know then that I would make my way to Norway to work with him in the years to come.

Then for my Ph.D., I was so lucky to be a research fellow at RITMO. Thank you, Anne, for all your effort as the director of this fantastic place. Realizing my doctoral project would not have been possible without the support of RITMO’s administrative team led by one and only Ancha. Thanks to the Research Council

of Norway for their generous financial support. I also want to thank the fourMs lab engineer, Kayla, and the lab engineer of the Robin group, Vegard, for their always promptly help in crisis times.

As a Ph.D. candidate, my "roller coaster ride" would be unbearable if some of my colleagues were not around. I want to thank Qichao, Agata, Victor, Tejaswinee, Marek, Ulf, George, Benedikte, Charles, Solveig, Olivier, Julian, Kjell-Andreas, Dongho, Mike, and everyone from the Interaction and Robotics Lab. I also should not forget to mention Kıvanç, my "academic comrade," who has been very close throughout this ride, albeit living in Canada's westernmost province. Neither should I forget the great people from the experimental music community in Oslo who welcomed me with open arms from the very beginning. Thank you, Morten, Kenneth, Fredrik, Eric, Thomas, Ingrid, and the HOTBOX crew.

With my utmost joy, I would like to dedicate this work to my parents, Arzu and Mesut, and my sister, Yağmur, who have taught and inspired me to be both a daring and a good-hearted human being. I have always felt their unconditional support next to me. Finally, my deepest gratitude goes to Katja; an inspiring artist and an amazing woman. Thank you for being a home to me, for all your inspiration, encouragement, intuition, and love. I am grateful to share life, create, and learn with you.

Çağrı Erdem
Oslo, June 2022

List of Papers

Paper I

Erdem, Ç., Schia, K. H., & Jensenius, A. R. (2019). Vrengt: A Shared Body–Machine Instrument for Music–Dance Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 186–191). UFRGS.

Paper II

Jensenius, A. R., & Erdem, Ç. (2022). Gestures in Ensemble Performance. In *Together in Music: Participation, Coordination, and Creativity in Ensembles* (pp. 109–118). Oxford University Press.

Paper III

Erdem, Ç., & Jensenius, A. R. (2020). RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 477–482). Birmingham City University.

Paper IV

Erdem, Ç., Lan, Q., & Jensenius, A. R. (2020). Exploring relationships between effort, motion, and sound in new musical instruments. *Human Technology* (pp. 310–347).

Paper V

Erdem, Ç., Wallace, B., Glette, K., & Jensenius, A. R. (2021). Tool or Actor? An Evaluation of a Musical AI "Toddler" with Two Expert Improvisers [Manuscript submitted for publication]. *Computer Music Journal*.

Paper VI

Krzyzaniak, M., Erdem, Ç., & Glette, K. (2022). What Makes Interactive Art Engaging? *Frontiers in Computer Science*, 4.

Contents

Chapter 1

Introduction

It don't mean a thing, if it ain't got that swing.
– Duke Ellington & Irving Mills (1931)

1.1 From Jazz Guitar to Performing with AI

I studied classical guitar at the conservatory as a teenager and moved on to study jazz, then computer music at the university. Playing the guitar or any other acoustic instrument was always an intimately physical and embodied experience. "You cannot swing if you don't dance," one of my guitar instructors once said. Jazz performers *feel* the groove and their body sway helps in creating and maintaining rhythm. As soon as I began performing music with computers, I started thinking about the lack of physicality. This made me wonder if it is possible to have an embodied engagement with computers similar to acoustic musicianship?

My curiosity led to the development of various wearable instruments,[1] e.g., shirts, spectacles, wigs, armbands and accessories. These instruments used inertial measurement units (IMUs) to capture the motion of body parts and various sensing technologies to measure muscle activity and other physiological processes. A Eureka moment came when I could abandon the guitar and just move in the 'air' to produce sound using these instruments. However, as it turned out, I became bored with the mappings between action and sound in these instruments. The fixed mappings I created were engaging at first but they did not change over time. To create more variation in my systems I began exploring the use of artificial intelligence (AI) techniques to overcome the problem. These reflections eventually led to many of this dissertation's questions regarding control and agency in interactive music performance.

[1] See http://cagrierdem.net/dev for an overview.

1.2 Motivations and Research Questions

The motivation behind the present dissertation is an urge to explore human and non-human entities controlling sound and music together, what I call *shared control*. I have explored such shared control in this dissertation through developing interactive systems based on four control strategies: (1) An instrument controlled by two human performers; (2) an "air instrument" with a chaotic control behavior; (3) an "air instrument" model based on the relationships between action and sound found in playing the guitar; (4) an audiovisual instrument controlled by a musician and a virtual agent. In Chapter 5, I will discuss the topics that emerged throughout the investigation, what has worked, where I failed, and what I believe should be done in the future. Ultimately, the goal is to contribute to the artistic, musicological, and technical understanding of AI through the lens of embodied music cognition. In particular, I have been interested in understanding more about how humans and non-human entities can share musical agency.

The main research objective of this dissertation is to:

> **Explore shared control between human performers and artificial agents in interactive performance to expand our understanding of agency and musical AI.**

Why would someone want to share performance control? My main drive came from the experience with limited controllability of 'air instruments.' For example, the muscle-sensing *Myo Armband* (see Figure 1.1), when worn one on each forearm, provides the user with 12 degrees of freedom (DoF). Here I am thinking about DoF as the number of independent motion variables in a mechanical system. But a count of 12 DoF does not necessarily give interactive freedom. When playing the electric guitar, one could argue that there are only 3 DoF—plectrum position, attack velocity, and finger position on the fretboard—and a relatively limited sound palette. On the other hand, computational sound-making possibilities are virtually endless. Even so, while playing the guitar, you can jump from one musical idea to the other in no time. This is difficult with an air instrument. First, in terms of precision and multitasking, moving a forearm in space is not comparable to having *hands-on* knobs, sliders, keys, or frets. Everything has to be set up before the show, and more parameters to control, e.g., changing presets, causes more cognitive load.

How is it possible to utilize low controllability as an interactive strategy? As a noise artist and musician, I was into practicing the aesthetics of indeterminacy and 'uncontrol.' Remembering John Cage's famous *the exploration of nonintention* (Cage, 1991), led me to ask about how machines could be given more initiative. The primary research objective became clear: *sharing the control* with musical agents. An analogy for that can be two persons playing the same guitar, one exciting the string while the other modifying the pitch on the fretboard. In technical terms, we can call these two persons as human agents. Agent comes from the Latin word *agere*, meaning "to do" (Russell, 2010). Such an agent does not have to be a living organism, hence be an artificial entity.

Figure 1.1: Myo armbands have been used in several of the projects presented in this dissertation. It is a (now discontinued) commercial sensor interface containing eight electromyogram (EMG) electrodes, an inertial measurement unit (IMU), and wireless communication over Bluetooth.

A *musical agent* is an artificial entity that can perceive, e.g., the performer, through sensors and act upon its environment by generating sound and displaying visuals. The perceptual inputs of the agent often called as *percept* are based on the physical signals, such as motion or audio. However, that is only a part of how we, as humans, move to make a sound. For example, while motion is an objectively continuous, uninterrupted signal, action is a segment in time where we aim to create a sound and move for that. As one can see, there are higher-level aspects of body movement. Then, how can musical agents interact with embodied entities, e.g., human performers? That brings me to the overarching research question:

How can embodied perspectives be included in developing musical agents for interactive performance?

The embodied perspective is concerned with an agent's percept for receiving an input and its processing abilities. More concretely, how can an *embodied perspective* in the agent program map percept sequences to action? To answer that, I draw on embodied music cognition theories and build upon a conceptual apparatus that defines movement at three levels (Paper II): *Motion* is a physics term representing the low-level signal domain. *Action* is mid-level and implies the psychological experience of motion to, e.g., make a sound or take a sip of a drink. *Gesture* denotes a high-level action with a a meaning-bearing component, such as swaying the head to signal the ending of a musical piece. I see all these levels necessary for designing perceptual monitoring systems of musical agents.

Throughout the dissertation, I will explore the interactions of human performers and musical agents from different perspectives: theoretical, empirical, and through design and performance. I am particularly interested in investigating interactive scenarios that use AI and multi-agent systems (MAS). Although there are examples of such systems in the literature, few of them have dealt with embodiment perspectives. The main research question can be broken down to three sub-questions:

RQ1: What are the relationships between action and sound in instrumental performance, and how can such relationships be used to create new interactive paradigms?

This question dates back to my experiences of performing guitars and various types of 'air' instruments. The latter allows for moving freely in space, which is liberating in many ways. However, playing on a guitar allows for using *force*, for example, while bending a string, jumping up and down with the dynamics and tempo, or playing a challenging part. Would it be possible to explore similar force-related control in an electronic system? Paper IV presents a statistical study and data collection, which deal explicitly with the action–sound relationships found in electric guitar performance, with a particular focus on measuring the muscle activity to estimate the force in sound-making. In doing so, Paper II provides theoretical support by clarifying the basic terminology of music-related body motion and drawing up some perspectives of how one can think about *gestures* in ensemble performance.

A second sub-question relates to musical AI and embodiment:

RQ2: What can AI offer for the action capabilities in interactive systems?

I here use *action capability* to imply a range of movement experiences led by specific goals, such as playing an E note on an empty string on the guitar. Paper I discusses a scenario where two performers control the same sonic and musical parameters. It does not use AI methods specifically but provides a conceptual model based on artists' feedback regarding shared control. Paper III builds on that concept and reports the evaluation of an air instrument performed in concerts with different ensembles. The air instrument used musical agents that automate interactive processes with ensemble members and generative algorithms that range from limited and constrained to highly open and surprising. Paper IV models sound-producing actions in playing the guitar to be used in the 'air' and compares different configurations of a particular deep learning method to map the muscle activity to sound. Finally, Paper V focuses on the evaluation of an interactive system, which used a generative deep learning architecture for an improvising virtual agent.

Reflecting on musical agents inevitably brings up the topic of *agency* and the third sub-question:

RQ3: What is the meaning of agency in interactive contexts?

Musical agency is here used as a *capacity to act* (Russell, 2010). Each of the developed systems have explored control strategies with different levels of complexity. Previous questions concerned the data and implementation, while this one is with the inquiry into performers' experiences regarding their communication with agents and expectations from an agency. Paper I investigates interactions between two human performers in a shared system and how they

dealt with control and the lack thereof. Paper V discusses two expert musicians' feedback about improvising with a musical AI "toddler." Paper VI reports the results of an online study, which investigated the varying level of agency that users ascribed to an interactive widget.

1.3 Scope

This dissertation encompasses a number of disciplines. My starting point was music performance or, even more broadly, the performing arts. Then I merged this with technology and got into mapping human body motion to sound. This inevitably led to focusing more on embodiment and approaches to bring embodied perspectives into the music technologies I use and develop. This dissertation, therefore, bridges from art to science in the attempt to explore the collaborative control of humans and machines in music performance. My creative work is inseparable from the analytic. Design and development have gone hand in hand. I have performed with my own systems but also evaluated others through both quantitative and qualitative studies. As such, the scope of the dissertation can be sketched as a mesh of relevant disciplines as depicted in Figure 1.2.

My methodological approach can perhaps best be described as an *iterative design process*. This has included moving fluently between conducting an extensive literature review, development of interactive systems, laboratory experiments, data collection, and observational studies.

Throughout the work, I have focused on three levels of music-related body movement concepts: At the "lowest" level I collected motion capture and physiological measurements and conducted statistical analyses about the relationships of motion and sound data. At the "middle" level, I emphasised the experience of sound-making on traditional musical instruments, and how I could create such relationships in new instruments. This has been done through development of new systems and instruments and evaluation via self-reports and interviews. At the "highest" level, I have explored the communicative aspects of musical human–machine relationships, with a particular focus on the performer's varying sense of agency in a collaborative performance.

I have had to limit the project in many directions. First, I have focused on electroacoustic sound-based aesthetics, prioritizing human–computer interactions in the signal domain. When it comes to sensing, I focused on muscle-based interaction from the start, although I have also worked with some motion and sound features. In the world of musical AI, I had to select only a few artificial intelligence techniques.

Various challenges emerged due to the interdisciplinary nature of this dissertation. Establishing a common terminology and addressing the theoretical and methodological requirements of different disciplines, particularly in collaborative studies, were among the prominent ones. According to the Kuhn–MacIntyre thesis as presented by Holbrook (2013), interdisciplinary communication can only happen when one learns the language of another discipline as a "second-first" language. Overall, I can place my work within the

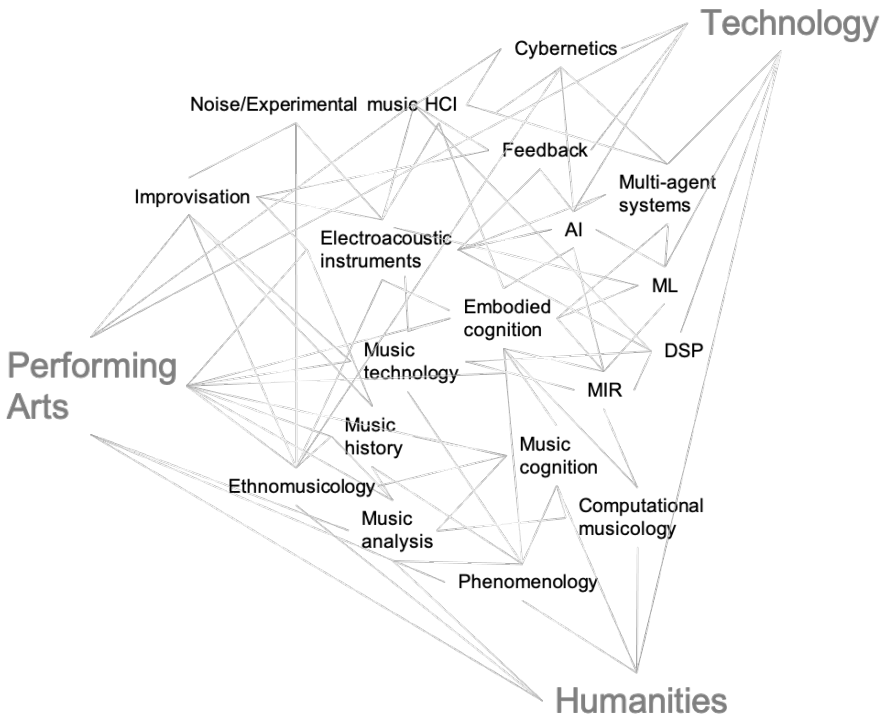


Figure 1.2: The relevant disciplines and topics that provide background for the dissertation project can be depicted as a mesh.

broad field of *music technology*, which is an inherently interdisciplinary research field that spans a wide range of disciplines, such as performing arts, musicology, human–computer interaction, and artificial intelligence (Serra, 2005).

1.4 Contributions

The present dissertation is genuinely interdisciplinary, and it contributes in different ways to knowledge development in all the disciplines that it encompasses. The main contributions can be summarized as:

- A literature review connecting perspectives from experimental arts, artificial intelligence, and musical embodiment.
- An empirical study of the sound-producing actions of guitarists that resulted in a multimodal dataset and a machine learning model.
- The iterative development of four interactive music systems, all of which are documented and made available for others to explore further.

- A number of artistic works put on stage in public events in the form of concerts (both physical and live-streamed), installations, radio broadcasts, and a released music album.

Inspired by the current global transition to Open Research practices, I have been careful to document all steps of the process and have made code and datasets.[2] That makes it possible for others to verify and replicate my work but also to use the generated material in new research and artistic activities. Just in the same way that I have benefited from the work of others, I hope others can build on my contributions in the future.

1.5 Thesis Outline

This thesis comprises two parts. The first part introduces the research motivation, the theoretical and empirical background (Chapter 2), and the methodology (Chapter 3) followed by the summary and discussion of the research contribution (Chapters 4 and 5). Figure 1.3 illustrates the intertwined connections between sections and chapters of the first part. The second part of the dissertation is a collection of six research papers that have been published or submitted to peer-reviewed journals and conference proceedings.

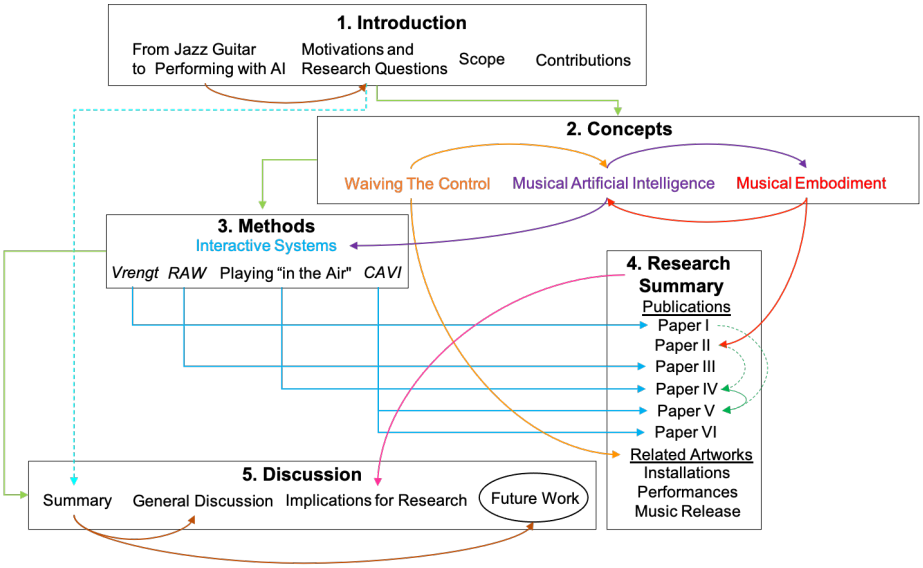


Figure 1.3: Chapter plan for this dissertation illustrating the pathways between sections.

[2]See Appendices A for links to repositories.

Chapter 2

Concepts

> *I believe that the use of noise to make music will continue and increase until we reach a music produced through the use of electrical instruments which will make available for musical purposes any and all sounds that can be heard. Photoelectric, film and mechanical mediums for the synthetic production of music will be explored.*
> – *John Cage* (1930)

This dissertation reflects a truly interdisciplinary journey; an artistic urge that has led to conceptual questioning, technical development, and performance exploration. It is in many ways a practice-led project, but with a basic research perspective. The aim has been to combine theories and methods from performing arts, computer science, and music cognition. A controlled experiment and user evaluation has helped create an 'objective' distance to the material. At the same time, many decisions were based on subjective experiences of my own performance practice. The physicality of sound, being on the stage, instincts and impulses, and bodily sensations all imbued an entwined research-creation. It can be summarized as a chase after an *unconventional* musical expression, a notion which I still find challenging to explain.

In this theory chapter, I will introduce different central concepts of the project. I have grouped them in three parts of which the first is called 'waiving the control.' Here I present an overview of my musical and aesthetical background in experimental music practice. This is important to understand where the project is coming from. Then follows a section on musical artificial intelligence (AI), its history, and relevant concepts, such as machine learning (ML), artificial agents, and a concept that emerged throughout my research: agency. The chapter closes with a discussion of musical embodiment. Here the aim is to reflect on AI from an embodied perspective.

2.1 Waiving The Control

2.1.1 Introduction

My interest in developing and performing with unconventional interactive systems that focus on indeterminacy resonates with Earle Brown's reflections on art (Nyman, 1999, p. 56):

> *What interests me is to find the degree of conditioning (of conception, of notation, of realization) which will balance the work between the points of control and non-control... There is no final solution to this*

paradox... which is why art is.
– Earle Brown

In my case, this pursuit has taken shape throughout years of extensively performing free improvisation, noise, and experimental music using various do-it-yourself (DIY) acoustic, analog, and digital electroacoustic instruments.

Noise and Control

It is common for experimental musicians to hack electronic hardware, such as household electrical appliances (Collins & Lonergan, 2020). One can also deliberately misuse products. For example, a mixing board can be transformed into an instrument by plugging the output to the input to make it self-oscillate (Figure 2.1). Such a *no-input mixing board* (NIMB), is well-known among noise and experimental artists. The principle is the same as creating acoustic feedback loops between a speaker and a microphone. Although there are some rare examples of meticulously controlled performances, such as Marko Ciciliani's composition *Mask* (2001),[1] NIMB is known for its emergent peculiarities (Charrieras & Hochherz, 2016). As a performer of such a system, your action capabilities are concerned with sharing musical initiatives with the tool, thus becoming less dominating and more dependent on the artifact.[2] In an interview, a leading NIMB practitioner Toshimaru Nakamura states (Paul, 2009):

> The no-input mixer is based on feedback. How can I explain... It's like sculpture. You shape the feedback into music. It's very hard to control it. The slightest thing can change the sound. It's unpredictable and uncontrollable, which makes it challenging. But, in a sense, it's because of the challenges that I play it. I'm not interested in playing music that has no risk.

According to Locke (1959), there is a two-stage temporal sequence in performing actions. First, possibilities randomly blossom as if they are freely "coming to us." Then, in the next phase called *de-liberation*, we choose one action possibility. When we act, what was previously out of control is now a determined action. In playing instruments such as NIMB, the thought and action processes, hence the decision-making, are distributed between the player and the tool's internal dynamics.

Poincaré (1914, p. 58) describes the process where his mind generates almost entirely out of his control as "[a] sudden illumination after a somewhat prolonged period of unconscious work." We can think of waiving performance control as a *two-stage model of free will*. In the first "free" stage, the system—instead of one's individual-self—generates the alternative possibilities within a certain range. Then, in the second stage, the performer uses her "will" to act; thereby,

[1] Video available at https://youtu.be/CoYE4QOWl3I.

[2] Audio recording of one of the author's solo improvisations on a no-input mixer is available at https://soundcloud.com/cagri-erdem/embodying071215.

Figure 2.1: A no-input mixer setup seen in the artwork for the cover of the album *Noise Mixer* by the noise artist *Don't Think*. (Think, 2019)

new forms of interactions and sonic outcomes emerge. The main difference here is the speed or amount of time for validating the options.

A Systemic Reading

We can develop an analogy between playing music and driving a boat. The helm of a boat—the space from which it is navigated—can be seen as analogous to the control interface of a musical instrument such as the knobs and faders of NIMB, mentioned above. The sea is the electrical current circulating in the components and becoming sound waves through the speakers. As the pilot, you look at where you want to go and regulate your boat's floating in that direction. You shift the steering according to the feedback from the environment concerning the waves, winds, and so on. In other words, you continuously evaluate the possibilities, introduce a move, and then validate the result before restarting the "loop." That is the basic understanding of *cybernetics*, which comes from the Greek word *kubernetes*, meaning the helmsman. Wiener (1948) was a central figure in the creation of what is now called *control theory*. He defined cybernetics as the entire field of control and communication theory, whether in the machine or the animal. In his thinking, the control problem is centered around monitoring the results of own operations, such as in homeostatic control mechanisms.

In the following, I will focus on the systemic aspects of performing music. Consider an example of music improvisation where you dominate the musical flow by playing too much or too loud in a live set. Then your improvisation partner may want to "pull you back" to a level that is more in line with the rest of the ensemble. The improvisation partner could also choose to follow, and continue to build on your uplifting riff. Both of these cases are likely to happen in collaborative improvisation. Such an improvisation can be seen as a recursive

system or mechanism that manifests adaptive behaviors within the environment. The improvisation actors' energy influxes imbue the system with negative and positive feedback. New information dynamically emerges, affecting progressive changes. There is a mutual interdependency that, particularly in electroacoustic practices, involves the individuals' embodied actions and various human-machine dynamics (Borgo, 2002; Borgo & Kaiser, 2010). "Feedback is the manifestation of interaction," argues Fellgett (1988). In this context, cybernetics can be seen as the science of interaction.

Control versus Configuration

Can what François (1999) called a "new cybernetic viewpoint" enable the production or an understanding of new forms of artistic consciousness? Artist–scholars, such as Borgo & Kaiser (2010) and Donnarumma (2016) suggest a mutual *configuration* with the (technological) practice. If your microphone faces the speaker too closely on a concert stage, thereby creating audible acoustic feedback, you will most likely be triggered to change the microphone direction spontaneously. This could be seen as similar to reaching out the hands while falling. In music, such spontaneity can be based on proprioceptive relationships between a musician and instrument (Paine, 2009).

All living systems are equipped with information-feedback paths to adapt to their environment (Kline, 2015). According to Maturana & Varela (1980), that is due to a particular character of the living systems: the *autopoietic* organization. Auto means "self," and poiesis, "creation" in Greek, hence autopoietic systems are ones that are comprised of self-creating processes (Straussfogel & von Schilling, 2009). In living systems, that refers to the circular (recursive) interactions between organisms' components (e.g., proteins, nucleic acids, lipids, etc.). For example, an individual living cell is:

> [...] a network of reactions which produce molecules such that (i) through their interaction generate and participate recursively in the same network of reaction which produced them, and (ii) realize the cell as a material unity. (Varela et al., 1974, p. 188)

From there, a broadened definition of an autopoietic system is:

> [...] a network of processes of production (transformation and destruction) of components that produces the components that: (i) through their interactions and transformations continuously regenerate the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network. (Maturana, 1980, p. 79)

Even though the concept of autopoiesis was initially developed to differentiate the living from the non-living (Mingers, 1989), the phenomenon is quite general. Dixon (2017) argues from a broader perspective that these principles are also

found in interactive arts. A cybernetic-autopoietic reading can be a helpful apparatus to understand better some interaction paradigms, regardless of their technological or non-technological nature. In particular, doing this reading in tandem with the shift of the artistic vision towards the *process*, and its influence on the music and performance, can provide insights into the link between autopoietic behavior and artistic urges that flourished in the 20th century. This way of thinking also paves the way to today's interactive human–machine paradigms.

Tanaka & Donnarumma (2018) has used the term "coadaptation" to describe the body–machine interaction between a human performer and a machine system. Can we read the motivation behind some experimental works and interactive paradigms, if not all, as an urge for *configuration* and waiving the control? Such a reading aligns with the cybernetic artist Ascott (2002) arguing for liberating the interactive art from the "perfect object." One may ask what is common between free improvisation, early avant-garde composers, and today's new instruments and musical artificial intelligence (AI)? It is interesting to think about both experimental music and cybernetic-systemic theories as a reponse to the emergence of machines.

2.1.2 Music as Process

The composer Steve Reich proposed *process music* with the aim of letting the audience hear the gradual steps that the composer took throughout the process of composing (Schwarz, 1980, 1981). He states: "[w]hat I'm interested in is a compositional process and a sounding music that is the same thing." This he exemplifies with John Cage's chance operations and *serial music* and argues that these works lack an audible connection between the compositional process and the sounding music. In his music, Reich wants the audience to hear how the building blocks are developed.

I also favor a process-oriented approach to musicking, although I focus on the *moment* and not on the individual past of neither the creator nor the creation. I am concerned with the sounding qualities of music as in the musical phenomenology of Schaeffer (1966). I am also concerned with the musical processes of the interaction between control and noise and the anarchic self-regulation of decentralized feedback paths routed through humans and machines. Thus, I call it *music as process*: the process of the musical experience. This resonates well with Ascott (1968, p. 2), who wrote:

> As feedback between persons increases and communications become more rapid and precise, so the creative process no longer culminates in the art work, but extends beyond it deep into the life of each individual. Art is then determined not by the creativity of the artist alone, but by the creative behaviour his work induces in the spectator, and in society at large. Where art of the old order constituted a deterministic vision, so the art of our time tends towards the development of a *cybernetic vision*, in which feedback, dialogue and

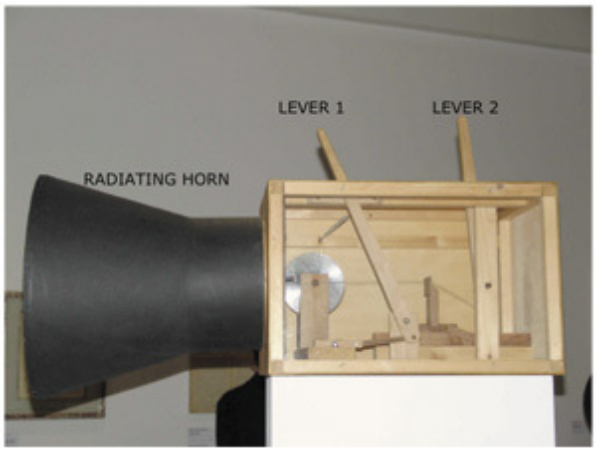


Figure 2.2: Inside of an Intonarumori re-built by Alexandra Spence for an exhibition. (Spence, 2016)

> involvement in some creative interplay at deep levels of experience are paramount.

Luigi Russolo's acoustic noise instrument *intonarumori* (Figure 2.2) was an embodiment of 20th-century industrial machinery. "This evolution of music is comparable to the multiplication of machines, which everywhere collaborate with man," states Russolo in his famous *Art of Noise*, critically pointing to Western music theory and its evolution since Greek antiquity (Russolo et al., 1913, p. 24). Russolo thereby foreshadows a fundamental paradigm shift regarding the notion of control in music with his *noise–sound* conception as "contention." In technical terms, noise denotes random (or stochastic) processes referring to irregular signal fluctuations. Russolo rejects a musical purity reserved to the refined expertise of composers, virtuosi, and luthiers. As a new way of music-making, he embraces the unpredictable nature of the noise; emphasizing contention as the taming of and adaptation to the uncontrollable.

Noise can be sound, and the lack thereof, or just "a purposeful purposelessness or a purposeless play" according to Cage (1961). His classic piece *4'33"* (1952) is a four minutes and thirty-three-second long composition consisting of three movements of "silence." Yet, as he clearly demonstrated, silence is not silent. In line with his lifelong "exploration of non-intention," Cage (1991) voids the musical content in traditional terms by waiving the control as composer. Instead, he lets the audience listen to sounds of the environment, sounds that are often ignored. In the way he uses the chance in his pieces, Cage seeks a balance between rational and irrational through random events within a controlled system (Jensen, 2009) and recommends a listening that focuses on constantly changing fluxes instead of trying to make sense of relations between sounds (Haskins, 2014). "Art can change our minds," Cage asserts (Cage & Goldberg, 1976) to stress our

Figure 2.3: A picture from one of the realizations of *Mikrophonie I*, performed by four players located around a tam-tam. (Muller, 2019)

minds' adaptive capacity. Thus, the intentionality that *4'33"* retains is to enact the audience to share musical agency and to create meanings centered around the parameters established by the composer (Cantrell, 2007).

Many new approaches to composition and performance emerged in the 1940s and 1950s. One piece that challenged the traditional understanding of musical control was Stockhausen's *Mikrophonie I* (1964). It relied on a single sound source, a large tam-tam, but the control was distributed between two percussionists and two more performers with hand-held microphones to amplify subtle details and noises (Figure 2.3). According to Burns (2002), a member of an ensemble that realized the piece, this new type of ensemble relationship was one of the fascinating aspects of the work. He describes the structure of *Mikrophonie I* as a "radical interdependence." No single player has complete authority over a particular sound event. Also, Stockhausen himself did not strictly score the piece and he left various compositional decisions to the performers.

The free jazz movement was one of the indeterminate music styles that came around the 1950s (Kosowitz & Vickery, 2013), about which Bailey (1993, p. 70) remarks: "passing over [the] control not to "chance" but to other musicians." Lewis (1996) takes that rather implicit critique to a more direct manner and argue that European avant-garde composers rejected the improvisational structure of jazz to preserve their genius as the composer despite all such new compositional approaches questioning the notion of control and authority. Fluxus was another contemporary that influenced artists to reconsider the isolated roles of the composer, performer, listener/viewer (Friedman et al., 2005; Magnusson, 2019). In George Brecht's *Incidental Music - Five Piano Pieces* (1961), for example, the performer keeps placing wooden blocks on top of another inside a grand piano until the blocks fall and excite the piano strings in an indeterminate

way.[3] As Ouzounian (2011) also points to, uncertainty was essential to Brecht's research. That sets a clear example of how artists started to question the notion of control in their works. Similarly, another piece, "*Bandoneon! (a combine)* uses no composing means since when activated it composes itself out of its own composite instrumental nature," wrote David Tudor in the program note to the premiere of his piece in 1966 (Goldman, 2012, p. 25). We can rightfully describe *Bandoneon! (a combine)* as a multimedia piece in today's terms. The "uncomposed" composition is a collaboration between Tudor, a video artist, a sound artist, and an engineer from Bell Telephone Laboratories (Rogalsky, 2010). The soloist instrument, the bandoneon, is a large concertina invented in mid-nineteenth-century Germany and migrated to South America (Goldman, 2012). It goes through a complex and unusual sound processing chain, including remote-controlled carts carrying speakers around the stage, a vochrome,[4] sound-reactive visuals, and various sound manipulations. *Bandoneon! (a combine)* is one of the first pieces that transformed the entire physical space into a self-oscillating instrument via acoustic feedback loops. "Once they are set in motion, they escalate like a forest fire," Goldman (2012, p. 54) describes the impact of the feedback loops.

2.1.3 Embodying Feedback

In a discussion of virtual bodies in cybernetics, Hayles (1999, p. 84) writes:

> Of all the implications that first-wave cybernetics conveyed, perhaps non was more disturbing and potentially revolutionary than the idea that the boundaries of the human subject are constructed rather than given. Conceptualizing control, communication, and information as an integrated system, cybernetics radically changed how boundaries were conceived.

A striking example of questioning the boundaries of the body and its integrity within the performance context is *Rhythm 0* by Marina Abramović (1974). The work involves Abramović standing still for around six hours while the audience members were allowed to use any of the 72 objects that were laid out to be used on her. These props included honey, wine, rose, and feather, but also a pair of scissors and a loaded pistol. The instruction was simple as Abramović tells in an interview: "I am the object. You can do whatever you want with me. I will take the responsibility for six hours."[5] The audience's participation ranged from giving Abramović a rose, kissing, carrying her around, taking her clothes off, and making her bleed. These actions can be seen as a form of *embodied feedback*. People's impulses became Abramović's bodily experience, and Abramović's embodied

[3] A video of a realization of the piece by Ben Vautier in 1985 is available at https://youtu.be/0n9818oCbJo

[4] An interface invented by Homer Dudley in the 1960s. The design is based on the vocoder. However, Vochrome, instead of synthesizing the voice, is used to map the audio signal to control the lights and sound (Kieronski, 1966)

[5] A video of the interview is available at https://vimeo.com/71952791

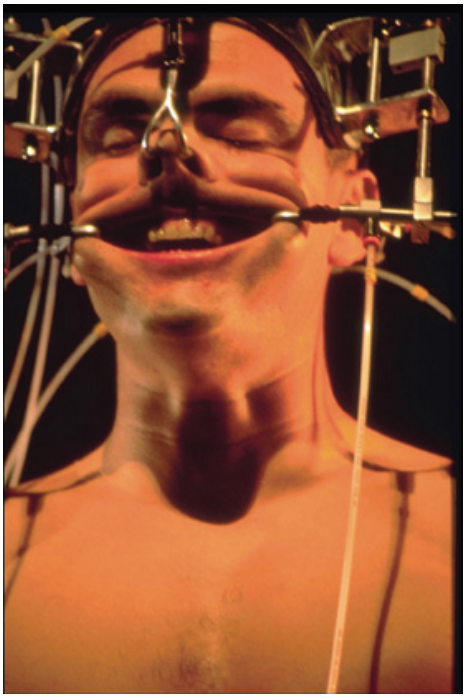


Figure 2.4: Marcel.lí Antúnez Roca performing *Epizoo*. (Photograph: F. Vargas) (Roca, 1994)

mind adapted to the experience. Ethical and aesthetical considerations aside, this performance is an example of non-technological mechanisms or paradigms that can affect progressive changes within an environment according to Dixon (2017).

In the late 1990s, the physical space that feedback can encompass extended immensely with the emergence of network technologies. In *Epizoo*, Marcel.lí Antúnez Roca exposes his body to the will of others (Roca, 1994). Via a video-game-like interface, the audience controls the robotic and pneumatic devices attached to different parts of his body while he is standing up on top of a podium (Figure: 2.4) (Jordà Puig, 2005; Donnarumma, 2016). In *Fractal Flesh* (1995), the performance artist Stelarc invites the audience to control his body using electrical muscle stimulation (EMS) (Figure: 2.5). According to Dixon (2019), this was a "cybernetic dance." In another example, as an Iraqi immigrant artist living in the United States, Wafaa Bilal escalates the narrative of the performance by emphasizing social and political aspects. In his *Shoot an Iraqi* (2007), he asks the audience to shoot him with a paintball gun controlled over the internet.[6]

[6]A video of a talk Wafaa Bilal focusing on his work *Shoot an Iraqi* is available at: https://youtu.be/WGhhYrHTGqI

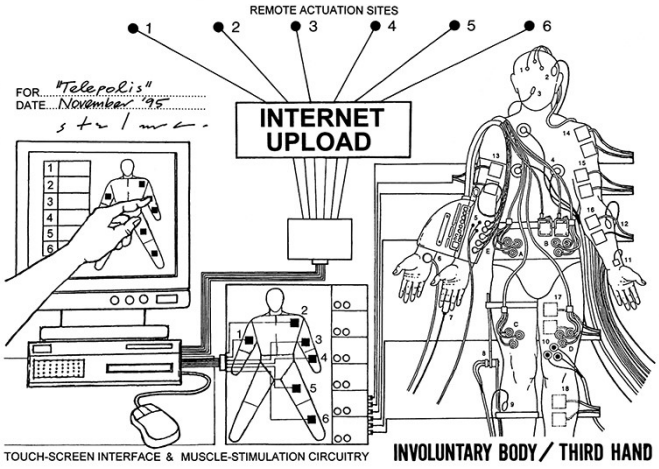


Figure 2.5: Stelarc, *Fractal Flesh*, performed at Telepolis, Luxembourg, 1995 (Stelarc, 1980)

The notion of feedback pervaded in the culture of the 1960s' North America. This included the daily language, such as in the question *may I ask your feedback?*. It also included popular music and experimental arts via the use of feedback instruments (Shanken et al., 2012). In the context of cybernetics, *biofeedback* emerged as a medical technique that uses electronic instruments to provide information about the physiological processes in humans or animals (Moss, 1999). The biodfeedback paradigm mainly incorporates monitoring techniques that provide feedback in the visual domain (visualization) and in the audio domain (sonification). From its inception, biofeedback focused on the control structures and autonomic functions within the organism. These include skeletal muscle responses that can be consciously controlled, or self-regulated processes, such as the heart rate (Anchor et al., 1982; Peper & Shaffer, 2018).

In the piece, *Music for Solo Performer – for enormously amplified brain waves and percussion* (1965) Alvin Lucier collaborated with Edmond Dewan, a scientist and a music enthusiast. Lucier writes about Dewan:

> He had been trying to interest Brandeis faculty composers in using his brain wave apparatus for musical purposes. No one was interested; perhaps they thought it was a gimmick. (Lucier, 2012)

The performance setup consisted of electroencephalography (EEG) electrodes placed on the performer's scalp (Figure 2.6). The electrodes captured the alpha rhythm of the brain (typically 8 to 12 Hz). These brain waves were first amplified through Dewan's "apparatus," then routed through an audio amplifier and mixer to 16 loudspeakers. The amplified alpha rhythms transform speaker cones into mechanisms that excite the sounding body of percussion instruments (Straebel & Thoben, 2014). According to Lucier's musical score, while the performer

Figure 2.6: John Cage (right) placing EEG electrodes on the scalp of Alvin Lucier (left) during preparations for *Music for Solo Performer* at the festival *John Cage at Wesleyan* (1988). Cage in particular encouraged Lucier strongly in the realization of this piece. (Rogalsky, 2010)

stands still, two assistants operate the 16-channel mixer (each channel routed to a speaker),[7] realizing the musical structure that the composer defined in advance. Two aspects of Music for Solo Performer are of interest to this dissertation. First, the human homeostatic system becomes a passive–active agent within a performative feedback loop. Second, the assistants somewhat embody the output from the performer's system and the composer–performer's instructions.

2.1.4 From Biofeedback to Biocontrol

Several works followed in the "feedback era" after Music for Solo Performer. John Cage's *Variations V* (1965) used several interactive processes involving dance (Miller, 2001). David Rosenboom's *Ecology of The Skin* (1970) also used brain waves (EEG) (Rosenboom, 1972). Stelarc used an electromyography-based (EMG) robotic body extension in 1980, the *Third Hand* (Dixon, 2019). Eventually, this biofeedback paradigm shifted into a new paradigm of *biocontrol* in the 1990s (Tanaka & Donnarumma, 2018). One of the first pieces here was Atau Tanaka's *Kagami*, featuring *The BioMuse* (Lusted & Knapp, 1988), which is a "biocontroller" that monitors the electrical activity in the body, in the form of both EMG and EEG (Tanaka, 1993).

Faster computers and interfaces allowed a widespread interest in using the human body as part of musical instruments at the turn of the 21st century. Particularly important here was the release of the commercial *Myo* sensor

[7] A video of the performance from 2010, featuring Alvin Lucier and two of his assistants is available at: https://vimeo.com/83631300

(see Figure 1.1), a wireless 8-channel EMG armband with a built-in inertial measurement unit (IMU) designed for human-computer interaction (HCI). The alpha version by *Thalmic labs* received 10,000 units of pre-orders in 2013. This gave numerous artists and researchers better access to exploring naturally occurring bioelectric signals in expressive audiovisual contexts. Since then, several custom software solutions have been developed to interface and process EMG and IMU signals (Kamkar, 2014; Françoise, 2015; Di Donato et al., 2018; Martin et al., 2018b), enabling numerous audiovisual applications (Benson et al., 2016; Jensenius et al., 2017; Di Donato & Dooley, 2017; Erdem, 2020). *Myogram* (2015) is a piece composed and performed[8] by Atau Tanaka using two Myo armbands on each forearm. In this piece, each of the (8) EMG channels is heard through direct audification routed to an octophonic sound system. The performer's overt body motion in the "air" is matched with the visceral peculiarity of the bioelectric muscle signals. Tanaka & Donnarumma (2018, p. 13) describes that experience as "spatial sound trajectories of neuron spikes projected in the height and depth of the space, with lateral space divided in the symmetry of the body."

Muscle contractions produce bioelectric signals but also mechanical vibration known as muscle *twitch*. Such twitches can be captured as acoustic signals through *mechanomyograms* (MMG) (Caramiaux et al., 2015). Donnarumma (2011) coined the term "biophysical music" to describe the intimacy of performing with such muscle signals in his custom device *Xth Sense*. This hardware uses an electret microphone-based armband to capture "muscle sounds" that are further processed via a custom Pure Data (Pd) patch. Donnarumma describes the piece *Ominus* "a relationship of configuration, where specific properties of the performer's body and those of the instrument are interlaced, reciprocally affecting one another" (Tanaka & Donnarumma, 2018, p. 15). Drawing on the design of the MMG sensor in Xth Sense, I developed *Biostomp*, a muscle-controlled motorized apparatus to be attached on the knobs of stompbox effects pedals for controlling the effects parameters using muscle contractions (Erdem et al., 2017).

The use of muscle signals spans robotic control to multimodal movement analysis. Françoise et al. (2014) investigated the use of interactive sound feedback for dance pedagogy based on Laban Movement Analysis (LMA). Ward et al. (2016) explored EMG and corresponding Effort qualities according to LMA, such as *flow*, being *free* or *bound* in dance. Fdili Alaoui et al. (2013) investigated expert movement knowledge for embodied interaction design via a thorough user-centered approach in collaboration with dancers and LMA experts. In their work, Niewiadomski et al. (2017) collected a multimodal dataset of dancers and studied the *lightness* and *fragility* qualities of expressive movement. Similarly, Sarasúa et al. (2017) investigated violinists' phrasings, using EMG and IMU signals captured by the *Myo* armband. Michailidis et al. (2018) explored building feedback paths, which can be seen as positively unusual—or "indirect," with their words—considering the open-loop design trend of new interfaces for musical

[8] A video is available at:https://youtu.be/G6H1J2k--5I

expression (NIMEs). They did so by mapping the signals captured via the Myo armband worn by a dancer to a haptic device worn by a pianist.

2.1.5 Uncertainty and Surprise

Over the years, I have performed with several different muscle interfaces. This includes the MMG- and EMG-based devices I have developed myself. I have also used various commercial products, such as the consumer-grade Myo armband and the medical-grade Delsys Trigno system. Based on all this experience, I am quite confident in saying that it is challenging to use muscle signals for precise control. I agree with Tanaka (2000) describing biosignals as "truly living signals." For example, it is challenging to maintain a stable tension level due to fatigue. The spectrotemporal resolution of sensor devices can be aberrant in the flow of action–sound causality. Moreover, from an artistic point of view, we can exploit the stochastic and non-stationary characteristics of muscle signals (Phinyomark et al., 2020) to procure a rich musical material (Ortiz et al., 2011).

We unconsciously execute several physiological and biological processes for a single, deliberate task (Chi et al., 2000). In other words, most human movement is found in the span between the conscious and the unconscious. In that regard, the biological signals produced by muscles reflect the in-betweenness of the human body's voluntary and autonomic functions. When we move as part of an action directed with a specific goal, the causality flows in one direction. Simultaneously, the dynamic interaction with the environment bestowing the body can flow back via the body's autonomic responses. In other words, the bodily experience of the environment feeds back into one's actions. That bi-directionality, no matter how fixed and controlled, gives rise to both uncertainty and surprise. The surprise element becomes apparent when trying to stand still for some time (Jensenius et al., 2017). I have experimented with such uncertainty and surprise in the software and pieces presented in this dissertation. In *Vrengt*, we experienced how EMG dynamics can vary while standing as still as possible, creating uncertainty that, according to the dancer, enabled "a new kind of body" (Erdem et al., 2019). Then in *RAW*, I built the entire musical strategy on sustained muscle contraction combined with uncertainty in the sound generation and real-time audio analysis to automate feedback from the other musicians in the ensemble (Erdem & Jensenius, 2020).

Uncertainty can be defined as a cognitive state based on a discrepancy between the desired information and the quality of the acquired (Ramirez et al., 2002). Unexpected events create prediction errors between the anticipation and the incoming input of the brain, which are often negative in valence. While uncertainty can lead to unpleasant experiences in real-life situations, such negative emotions are arguably an essential resource for positive aesthetic experiences (Menninghaus et al., 2017). Ludden et al. (2008), for example, proposes to design *sensory incongruity* between how a consumer product looks and feels to motivate people for further exploration. Unexpected events are more alerting and likely to be remembered compared to predictable ones (Ranganath & Rainer, 2003). Shany et al. (2019) suggest that surprising elements in music

create auditory 'alerts' leading to raised arousal levels. That is in line with studies linking the pleasure of music-listening to uncertainty and surprise in popular music (Huron, 2019) and in the atonal music composed over the last century (Mencke et al., 2019).

John Cage's *chance operations* used random events to function within the context of a controlled system (Jensen, 2009). He describes his focus as "exploring non-intention" and sees the use of randomness and noise as a liberation from the artist's intent and emotions and a mimickry of nature's indeterminacy (Cantrell, 2007). David Borgo, an improviser and scholar, links free improvisation to the complex and unpredictable dynamic systems found in nature (Borgo, 2005). In the context of interactive music systems, Chadabe (2002) argued that one-directional and deterministic action–sound mappings underutilize the means of computers to provide the human performer with complete control. More recently, uncertain and surprising processes have been explored through a variety of concepts and methods: The "uncontrol" paradigm using Echo State Networks (Kiefer, 2014), feedback instruments (Liontiris, 2018; Melbye & Ulfarsson, 2020), emerging dynamics via distributed control in mechatronic instruments (Gurevich, 2014), creative inaccuracies of artificial neural networks (Snyder & Ryan, 2014), breakpoints of machine learning (ML) algorithms for creative unpredictability (Schacher et al., 2015), and nonlinear dynamical processes (Berdahl et al., 2018; Mudd et al., 2019).

2.1.6 Summary

I started this section by introducing the use of feedback instruments in noise and experimental music. That was a brief yet critical reflection on how I approach sound and music-making. Then I presented and discussed cybernetics and how a cybernetic perspective can be connected to what I call *music as process*. In the 20th century, avant-garde composers waived the authorship of their music, leaving many or sometimes all the outcomes to chance. In a discussion on *embodying feedback*, I presented works that depict human bodies deliberately used within feedback loops at varying abstract and physical senses. While performance artists left themselves to the mercy of spectators, letting the audience do all sorts of things to them, others used biological processes in sound-making. Then the bodies can be thought of as generative "machines," that can be used to discover and wander around several control paradigms. Ultimately, this boils down to a critical topic in this dissertation: *uncertainty and surprise* in interactive music.

2.2 Musical Artificial Intelligence

2.2.1 Introduction

Although AI has indisputably become one of the hottest topics in the last decade, the idea of automated artifacts dates at least as far back as the beginning of humankind's written record. The first premise of today's rule-based systems based on the *if...then* condition can be found in *modus ponens* of antiquity.

To our knowledge, Ctesibius' water clock or *clepsydra* (c. 250 BC) is the first machine ever built that can operate under its own control. The device could automatically empty the reservoir through a siphon and reset itself. Such a systematization of *modus ponens*, which is quite essential for modern computing, created the basis of the first artifact that can adapt its behavior to changes in the environment. Self-regulating feedback control, which was once only a feature of living systems, has become the focal point of modern machines. The word *algorithm* goes back to Persian mathematician al-Khawarazmi from 9$^{\text{th}}$ century. One of the first AI programs, Newell and Simon's implementation of *The General Problem Solver (GPS)* (1959) is based on an algorithm[9] suggested by Aristotle in *The Nicomachean Ethics* (Newell et al., 1959). Thomas Hobbes, who later became one of the "prophets" of AI, described his imagery of "artificial animal" in the 17$^{\text{th}}$ century (Hobbes, 2001), around the time when the first mechanical calculators were built by Blaise Pascal, then by Gottfried Wilhelm Leibniz (Russell, 2010). Until the AI's inception in the 1950s, various scholars and thinkers from different fields, such as philosophy, mathematics, economics, neuroscience, and psychology, mused about the mechanization of human thought and action in non-human entities, eventually giving rise to artificially intelligent agents. In this section, I will embark on a journey starting from the musical automata of prehistory into the musical applications of AI technologies in the 20$^{\text{th}}$ century. Then, I will give a brief overview of the state-of-the-art techniques and their use in music before jumping into the topic of musical agents and agency.

2.2.2 History

Musical Automata

Based on the preserved historical documents, such as the Arabic translation of Archimedes' treatise (Apollonius, 250BC), we can trace, e.g., Archimedes' and Apollonius of Perga's flute-playing automata, back to the 3$^{\text{rd}}$ century BCE (Krzyzaniak, 2016). The treatise that Hero of Alexandria wrote in the 1$^{\text{st}}$ century AD, *Pneumatica*, contains several sound-making automata, such as singing birds. In the 9$^{\text{th}}$ century, The Banu Musa brothers developed an automatic flute player using water and air pressure (Farmer, 1931). In the Book of Knowledge of Ingenious Mechanical Devices published in 1206, Al-Jazari gives the details of an automated water-powered percussion orchestra on a floating boat designed for entertainment at royal drinking parties (Figure 2.7) (Hill, 1974). Al'Jazari's musical boat is often considered the first programmable robot. Around that time, Dutch engineers developed binary programmable carillons (Buchner, 1978). Marie-Dominique-Joseph Engramelle presented plans of cylinder-driven instruments (Engramelle, 1775), which could be applied to both mechanical musical instruments and automata in the form of human beings or animals (Kemper & Cypess, 2019). In the 18$^{\text{th}}$ century, Jacques de Vaucanson

[9]GPS could solve the tricky missionaries-and-cannibals puzzle, which requires one to go backward to go forwards (Boden, 2006, p. 324).

Figure 2.7: The musical boat of Al-Jazari, Topkapı manuscript, 1206. It consisted of musical automata built to entertain the guests at drinking parties at the King Court in Diyarbakır, located in the southeastern Anatolia region. (Golan, 2019)

invented several automata, perhaps the most famous of which was a flute player (Vaucanson, 2018). Several remarkable automata, such as Mareppe's violinist, Manzetti's flutist (Figure 2.8a) and Kaufman's trumpetist (Figure 2.8) automata appeared during the 19$^{\text{th}}$ century. Although there are also later examples of automata, one can say that this mechanical age ended with Edison's invention of the phonograph and the microphone by Berliner in 1877.

The Electronic Age

In his treatise from 1948, *The Mathematical Basis of the Arts*, Joseph Schillinger argues that any art form is a measurable quantity as it is manifested through a physical medium and perceived through a sensory organ, hence can be automated and systematized. He created several algorithmic techniques that, for example, produce various linear designs utilizing rhythmic series, using angles, dimensions, directions, and their derivatives (Schillinger, 1948, p. 363). In 1951, computers generated musical melodies for the first time in Alan Turing's Computing Machine Laboratory at Manchester University (Copeland & Long, 2017). In 1956, Nicolas Schöffer created *CYSP 0 & 1*, which are human-scale robotic sculptures that can respond to changes in the environment, such as sound, light intensity and color, and movement (Shanken et al., 2012). Schöffer remarks about the paradigm shift in the creation of arts in the electronic age (Whitelaw, 2004, p. 17):

> We are no longer creating a work, we are creating creation. [...] We are able to bring forth...results...which go beyond the intentions of

(a)

(b)

Figure 2.8: (a) Manzetti's flute automaton in the Saint-Bénin exposition center. (b) The Kaufmann Trumpeter. (Hoggett, 2012a,b)

their originators, and this in infinite number.

That resonates well with the term *generative aesthetics* that Max Bense defines as "the artificial production of probabilities of innovation or deviation from the norm." Bense coined that term in an article entitled *projekte generativer ästhetik* or the projects of generative aesthetics (Nees & Bense, 1965), which was published as part of the first-ever generative art exhibition,[10] *Generative Computergraphik*, in 1965. According to Nake (2012), this text can be seen as the manifesto of algorithmic art.

The influence of industrial machinery on new approaches to sound- and music-making at the beginning of the 20$^{\text{th}}$ century (Section 2.1.2) soon witnessed a boom of post-war technologies, emerging theories, and their penetration in the art forms. The proliferation of magnetic tape for music production in the late 1940s and the deployment of the first computers in the 1950s were pivotal. The first AI computer program that solved non-numerical problems, *Logic Theorist*, was co-authored by Herbert Simon, Cliff Shaw and Al Newell in 1956 (Russell, 2010). Simon (1996, p. 190) later claimed in is autobiography that the program "solved the venerable mind/body problem." Pierre Schaeffer and colleagues established the *Groupe de Recherche de Musique Concrète* (GRMC) in France in 1951, and several composers, such as Edgard Varèse, Olivier Messiaen, Pierre Boulez and Iannis Xenakis worked there. Around the same time, Werner Meyer-Eppler led the Cologne Studio, where Stockhausen, Luigi Nono, and John Cage produced notable works (Collins et al., 2013). By the 1960s, the interest in computer-generated art escalated, spanning an artistic-scientific interdisciplinarity. An important indicator is the *Cybernetic Serendipity* exhibition, which happened in

[10] According to Nake (2012), it was Joan Shogren who made a screening of computer-generated drawings for the first time in 1963 at a university.

1968 with 130 contributors, of whom 43 were composers, artists, and poets, and 87 were engineers, doctors, computer scientists, and philosophers (Reichardt, 1968).

Early Research on AI & Music

Over the past centuries, many composers employed *Würfelspiel* combinatorial techniques to create musical works (Roads, 1980; Cope, 2001). It is also possible to see exciting crossing points between emerging algorithmic approaches and ancient collaborative music practices. An example of which is the *Change Ringing* from the early 17^{th} century England. In this practice, there is n number of bells, each controlled by a single person. Hence, a group of n number of persons ring the bells one by one, every round choosing one particular order among $n!$ amount of possibilities; the aim is to perform all the permutations (Strickland, 2018). The flow of the physical coordination is crucial. In this case, each agent– a human– is part of an autopoietic organization internally and externally with the others. Hiller & Kumra (1979) conducted a research project nearly two centuries later, where they investigated the permutational technique of the Change Ringing for algorithmic music composition.

By the time the first attempts at computer-generated music appeared in the 1950s, focusing on algorithmic music creation, novel approaches to composition and performance were already germinating (see Section 2.1.2). Around the time when the interest in using audio feedback and tape loops was blossoming, the integration of computers in music was at its early steps. Iannis Xenakis led stochastic music by introducing the theory of probability in music composition. He explored probabilistic calculus in his large-orchestra works, first in *Metastasis* (1955), then, *Pithoprakta* (1957), before he started to use computers in 1960s for stochastic calculations (Serra, 1993). Also, the use of multi-agent systems (MAS) in composition and performance can be traced back to Xenakis' ideas of using game theory (Miranda, 2011, p. 166).

In 1957, *Illiac Suite* became the first composition that was entirely made with computational means (Hiller & Isaacson, 1979). Illiac is the abbreviation of *Illinois Automatic Computer*, which was one of the first computers ever built in the US (Figure 2.9). In doing so, Hiller (composer) and Isaacson (composer and mathematician) defined a set of rules to map the random numbers generated by *Monte Carlo* methods to musical features. In terms of Illiac Suite's place in the music history, Luc Steels points to how it was "composed at the time John Cage's experimental music had come in vogue, emphasizing aleatoric elements, processes, and rhythm and tone rather than melody and harmony" (Miranda, 2021, p. vi).

Following Sumner and Simon's formalization of patterns in tonal music and Winograd's harmony-analysis program in the 1960s, a breakthrough towards the so-called *musical AI* came in the late 1970s with the *generative modeling* of music. While algorithmic music aimed at creating "aesthetically satisfying new composition," generative modeling proposed new material based on the analysis of a corpus of compositions (Roads, 1980). Cope (1989) with his *Experiments in*

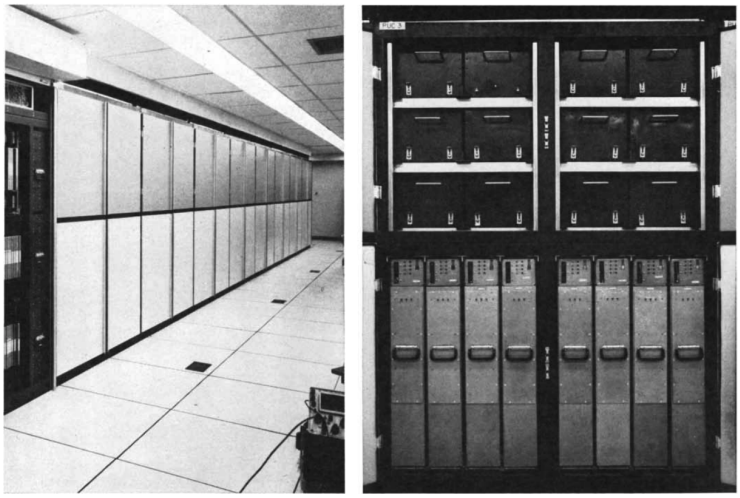


Figure 2.9: A picture of *Illiac*, which was of the first computers ever built with thousands of vacuum tubes, weighting around five tons. (Slotnick, 1971)

Musical Intelligence (EMI) project in 1983 not only explored music composition using language models with augmented transition network (ATN) but also suggested the concept of "musical intelligence." That is, the music and attributes are encoded, and musical segments are extracted using pattern matching before being categorized and reconstructed using ATN (Ranwala, 2020). EMI was able to generate hundreds of compositions based on multiple composers' works.[11] Cope's work had a notable influence on many subsequent musical machine learning (ML) and AI technologies.

2.2.3 Machine Learning as Tool

Machine learning (ML), as a subset of AI, aims at building mathematical models based on the given examples called "training data," to make predictions or decisions without being explicitly programmed to achieve the task. Dahlstedt (2021) discusses how artists can use such models and algorithms as tools at various stages of the creative process. This development has escalated with the rise of deep learning (DL) over the last decade. Deep learning is a subset of ML where artificial neural networks allow computers to understand complex phenomena by building a hierarchy of concepts out of simpler ones (Goodfellow et al., 2016). Face recognition, music recommendation, translation, and image classification are just a few applications of ML/DL in everyday life. It is fairly recent that these techniques have reached wide interest. However, ML has been an important component in the design of and performance with new interfaces for musical expression for decades (Lee et al., 1991).

[11] An audio demonstration of Bach-style Fugue 1 from the well-programmed Clavier by EMI is available at https://youtu.be/Lt7fEchgFrU

Zhang (2020) groups learning methods of machines into three categories: (1) the network structure, such as artificial neural networks (ANNs) or Bayesian network; (2) statistical analysis, which encompasses a wide range of algorithms and applications, such as clustering, Hidden Markov Models (HMMs) and Naive Bayes; and (3) evolutionary computation. Following the proliferation of faster computers and digital protocols, many easy-to-use tools have been developed over the years for artists and musicians. Fiebrink's *Wekinator* is an open-source platform that provides supervised learning algorithms to real-time problem domains, such as interactive computer music (Fiebrink, 2011). Among others, Gesture Follower (Bevilacqua et al., 2010), the SARC Eyesweb catalog (Gillian, 2011), ml.* library (Smith & Garnett, 2012), Gesture Recognition Toolkit (GRT) (Gillian & Paradiso, 2014), Gesture Variation Follower (GVF) (Caramiaux et al., 2014b) and ml.lib (Bullock & Momeni, 2015) allow the application of ML algorithms through either a graphical user interface (GUI), or, in the form of external libraries for audio programming platforms, such as Max/MSP and PureData (PD). We can group the main musical applications that employ ML under three broad categories: mapping, analysis, and generation. These will be discussed separately in the following sections.

Mapping

Action–sound mapping is critical in most electroacoustic instruments (Jensenius, 2007). Developing mappings is a way of creating relationships between the input of a system (e.g., sensors, buttons, faders, etc.) and a sound engine. Mappings can be created manually, but ML tools allow for creating complex relationships. Applying ML algorithms in mapping structures has become widespread with the tools and frameworks mentioned above. The most used type of ML for mapping is *supervised learning* (SL), in which the algorithm models the relationship between the input data and the labeled "target" output data (Fiebrink & Caramiaux, 2016). Since the early use of ANNs by Lee et al. (1991), SL has been used for the mapping between the performer's action and sound. SL is mainly used for *regression* (for continuous outputs) or *classification* (for discrete outputs) (Zhang, 2020). The former can be used, for example, to map continuous body motion to a sound feature (e.g., pitch), for which ANNs are a convenient modeling method. The latter, on the other hand, can be used for triggering or selecting actions. K-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Adaptive Boosting (AdaBoost), and Naive Bayes are some of the common classification algorithms.

Among other methods for mapping, *unsupervised learning* (UL) is used with unlabeled training data, where the algorithm learns the internal representations. Common UL tasks include clustering, classification, and dimensionality reduction (Smith & Garnett, 2012). In a study by Prpa et al. (2018), UL is used as the musical agent architecture that selects samples from the audio corpus according to the frequency of the user's breathing. *Reinforcement learning* (RL) is situated between SL and UL. In RL, the agent learns from a series of feedback—rewards or punishments—given by another human or algorithm (Wiering & van Otterlo, 2012). Visi & Tanaka (2021) used RL for exploring the mapping possibilities

between input sensor data streams and sound synthesis parameters, which they call "assisted interactive machine learning" (AIML).

Analysis

In the context of human–computer interaction analysis can be seen as "machine understanding" of a (musical) gesture (see Section 2.3.4 for an overview of terminology) by processing its features over time (Fiebrink & Caramiaux, 2016). Different from building a relationship between the input and target data as in *mapping*, here the analysis covers the real-time procedures that process and analyze musical gesture data to use as part of the system's control structures. The data to be analyzed can be a sensor signal, a symbolic musical instrument digital interface (MIDI) data, an image or video, or an audio stream. For example, a *machine listening* model can classify an input audio frame (Mishra et al., 2018; Purwins et al., 2019), or, an input of motion frame (Caramiaux et al., 2013; Côté-Allard et al., 2019). The output can trigger or adjust an internal process, which may not be explicitly associated with the input gesture. This process can be that of synchronization, alignment, or imitation. Therefore, I group such a broad range of applications under the title *analysis*, which is different from causing a perceivable outcome or generating new content.

From a music theoretical perspective, it is possible to calculate the probability distribution of interrelations between chords to predict what chord comes next. For example, in mainstream jazz, the probability of *E-7* & *A7* chords followed by a third-degree *F#-7* chord is much less than that of the first-degree chord, *Dmaj7*. Chord progressions proceed sequentially. In terms of what will come next, the present sequence has greater importance than the past one. That echoes well the Markov property, which states that the future depends only on the present and not on the past (Puterman, 1994). The term Markov property refers to the memoryless property of a stochastic (random) process. It is assumed that we can estimate the parameters of "what comes next" by modeling stochastic processes as a sequence of states (Rabiner, 1989). Eventually, transitions between states converge to a specific probability, such as the likelihood of the two chords as mentioned earlier, and the states form a *Markov chain* (Figure 2.10). Among other statistical models (e.g., Gaussian and Poisson processes), the most common methods for statistical sequence modeling of music are Markov models (e.g., Variable Markov Models (VMM), Hidden Markov Models (HMM), etc.) and Factor Oracles (FOs). Markov models, in particular, have been used extensively in interactive music systems to track and learn from the ground up during performances in real-time.

Listening is a core modality of a musician, and many learning algorithms have primarily employed machine listening strategies. The tasks include detecting low-level features, such as beat onset, or high-level features for event analysis (Rowe, 1992; Collins, 2006). Cont (2010) explores an HMM framework for the real-time alignment of audio signals to symbolic music scores, called *score following*. The *Gesture Follower* system of Bevilacqua et al. (2010) allows the synchronization of digital media and effects to a given reference gesture. The GVF of Caramiaux

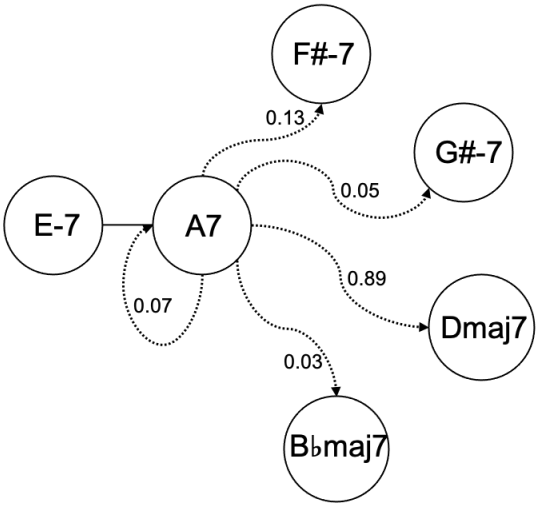


Figure 2.10: A simplified graph illustrating the weights of the possible next chords following a standard II-7 & V7 progression.

et al. (2014b), as used in an augmented piano system by Zandt-Escobar et al. (2014), provides continuous information about the speed, scale, and rotation of the expressive motion. Following the frameworks of Caramiaux et al. (2014a), Sarasua et al. (2016) proposed an electroacoustic instrument which adapts the mapping for each user by observing spontaneous conducting movements. In a similar project, one classifier recognizes a conductor's gestures by describing the musical features (e.g., pointillistic, long tones, noise, etc.) and signals section beginnings and endings to other musicians in the ensemble. Another classifier recognizes the finger gestures pointing to a particular player in the ensemble (Nort, 2018). As Purwins et al. (2019) report, emerging DL methods have become state-of-the-art for audio signal processing of speech, music, and environmental sounds. These new DL methods can provide more complex mappings and analyses. However, they also have much higher computational needs than, for example, Markov models. This causes some challenges when it comes to latency in real-time settings.

Generation

The *Continuator* by Pachet (2003) is a well-known example of a generative music system. It can autonomously continue the musical sequences played by a human performer, based on the kind of style of the training dataset. The Continuator tracks the performance through MIDI, and as soon as the performer stops playing, it starts generating temporal sequences modeled using a variable-order Markov model. While that model is trained on the performer's playing, a more recent similar system, the *AI Duet*, uses a deep recurrent neural network (RNN) trained on a large dataset to predict new notes that are likely to come

next (Mann, 2016). Different from former *feedforward* networks, RNN introduces the concept of memory (Haviv et al., 2019). One popular model is the Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997). In addition to the easy-to-use frameworks listed in Section 2.2.3, Martin & Torresen (2019) has created a framework that focuses on the use of RNN-LSTMs in interactive music performance. He stresses the concept of *proactivity* of musical instruments by employing deep RNNs to make creative predictions in return to the human performer's input. I have used LSTMs for audio RMS prediction based on muscle signals in "air guitar" context (Paper IV). LSTMs can also be used generate full-body dance motion based on audio features (Wallace et al., 2020).

For decades, musical AI has dealt with the fascinating problem of generating music in the symbolic domain (e.g., MIDI). Some recent examples include *folk-RNN* (Sturm et al., 2015), *Impro–Visor* (Johnson et al., 2017), and *DeepBach* (Hadjeres et al., 2017) (see Briot & Pachet (2018) for an overview). While the research on the symbolic representation and (re)construction of music is still a nontrivial problem, the arrival of massive autoregressive DL models introduced the sub-symbolic exploration of music generation in the waveform domain. These models can use a convolutional neural network (CNN), such as *WaveNet* (Oord et al., 2016), a stack of recurrent neural networks (RNNs), such as *SampleRNN* (Mehri et al., 2017). Engel et al. (2017) developed *NSynth* using the WaveNet autoencoders to explore neural audio synthesis of musical notes. The artist-scholar duo, *Dadabots*, trained the latter network, SampleRNN, with a dataset of audio tracks in metal and punk genres.[12] For their aesthetic choice, they remark (Carr & Zukowski, 2018, p. 2):

> Music genres like metal and punk seem to work better, perhaps because the strange artifacts of neural synthesis (noise, chaos, grotesque mutations of voice) are aesthetically pleasing in these styles.

Other audio generation frameworks include generative adversarial networks (GANs) (Engel et al., 2018; Donahue et al., 2019), variational autoencoders (VAEs) (Tatar et al., 2020) and combining traditional signal processing with neural networks (Engel et al., 2020). In a co-creative system aimed for sound design, Scurto et al. (2019) implements an interactive framework using deep RL agents that learn from interactions with humans. As one of the rare NIMEs that employ GAN-based audio synthesis procedure is the *AI-terity*, which is a non-rigid, bendable instrument that focuses on surprising and autonomous features (Tahiroglu et al., 2021). Most works and examples I have listed under the title *generation*, in general, and AI-terity, in particular, point to an interesting transition situated in between *ML as a tool* and *AI as an actor*.

[12] A 24-hour live stream of the model generating technical death metal can be found at: https://youtu.be/MwtVkPKx3RA

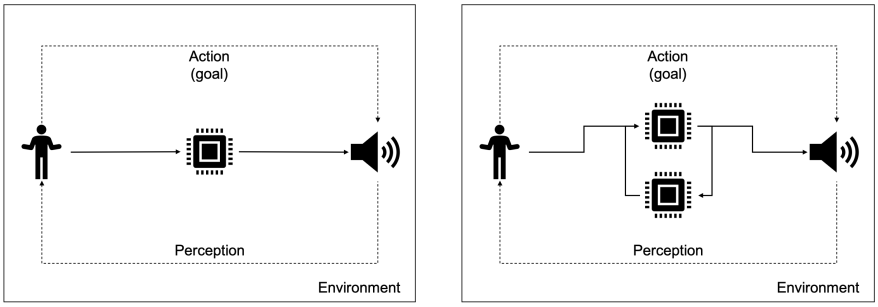
2.2.4 From Tool to Actor

In social sciences, such as in the Actor–Network Theory (ANT), mundane objects are argued as actors depending on their influence on the outcome (Latour, 2005, p. 153). For example, the gauge and material of a guitar string dramatically effects how it sounds. Hence, the material-instrument denotes an agency, which Mendoza & Thompson (2017) argue in close relation to the performer's "gestural agency." In computer science, the term *actor* was coined by Hewitt et al. (1973) as a unified formalism that denotes distributed agents functioning within a system. According to the *actor model*, concurrent computational processes, each suggesting an *actor*, are formalized as "special cases" of the agents that communicate by sending and receiving messages. The terms actor and agent are often used interchangeably. Both denote an entity that acts on behalf of another entity. In computer science, an actor is linked explicitly to a computational process. In systems theory, on the other hand, an actor can refer to any arbitrary system or one of its components, such as living beings, artificial systems, or imaginary characters (Burgin, 2017). That connects well to the artistic-performative contexts that I am interested in.

Caramiaux & Donnarumma (2021) distinguishes using ML algorithms for action–sound mapping from employing AI as a separate entity to perform together. In this regard, Dahlstedt (2018, p. 18) suggests three modes of performance in interactive systems: "Performing *on*, *with*, and *in* algorithms." Performing on algorithms refers to controlling the algorithm's parameters, as in most electroacoustic instruments. Performing with algorithms implies more autonomous processes of the algorithm, which can have some influence on the performer's actions. Finally, performing in algorithms denotes strong bi-directional feedback paths between the human and the machine where the performer becomes part of the system. Dahlstedt (2018) describes this as a *systemic improvisation*.

Nymoen et al. (2016) point to a sweet spot between acoustic instruments and media device/services (e.g., *Spotify*, etc). They propose instruments containing a self-awareness, that is, a level "beyond stimulus-awareness," such that it can allow what they call "active music" that is somewhere in between the high ceiling of traditional instruments and media systems' lack of controllability. According to their approach, the *tool* end of the continuum refers to "hard-earned skills" while the *actor* end refers to a lack of control. All in all, the difference between a *tool* and an *actor* is related to the causality flows within the interactive scenario. This also ultimately also boils down to how the system is perceived and experienced.

When it comes to perception, the dominant view is a cyclical process in which the brain triggers the sensory organs so that the information from the environment can flow through them. Thus, perception can be seen as a closed-loop system (Ahissar & Assa, 2016). In this model, sensory feedback (e.g., tactile) is the intrinsic source of control. NIMEs or interactive music systems are often intended as open feedback systems (Figure 2.11a). Such systems are often based on the causality between action and sound (Hunt & Wanderley, 2002; Jensenius, 2007; Van Nort et al., 2014) . In the case of an acoustic guitar, we are 100% sure



(a) Open feedback system. (b) Closed feedback system.

Figure 2.11: Feedback pathways in interactive music systems.

that the player makes the sound. The guitar abides laws of physics, and it is not made to surprise the skilled user. The player aims to play a note, acts, and the guitar transmits the physical features of the body motion into sound in the most transparent way, resulting in the player authoring the entire music-making process. The same has not been the case with digital musical instruments, in which the level of casualty between action and sound may differ.

Expert players of traditional[13] musical instruments develop a refined technique over years of practice. There is often a negative correlation between the instrumental skills and the opacity of the instrument. On could argue that, as more skills develop, the tool tends to become an extension of the human body. Similar to how we do not actively think about our body parts in daily movements, the instrument also, as an extension of the body, becomes transparent (Nijs et al., 2013). The performer's expertise and authorship are tied to the disappearance of the instrument. There, the tool differs from the actor. Some teachers in the conservatory high school used to tell us, "the music made with machines is not real." In their thinking, machines were interfering with our "acts" as musicians.

Consider a musician with a guitar. Then add an electronics effects pedal. Then another pedal, then another, and perhaps a computer too. Suddenly you arrive at a setup in which the human performs with a chain of devices that all make its own set of decisions. Who makes the sound in such a system? Who is an actor? To what extent and in what ways is the musical agency distributed among human and non-human entities? There is an unconventional in-betweenness of controlling the machine and being controlled. That echoes Schöffer's reflections on his computer-driven cybernetic art: "We are no longer creating a work, we are creating creation" (Whitelaw, 2004, p. 17). For what Schöffer signaled in the 1950s, artificial agents have become a significant modeling paradigm towards the end of the 20$^{\text{th}}$ century.

[13]Here, the term, traditional, refers to what Smalley (1997) explained as an intuitive knowledge of action–sound causalities in traditional sound-making.

2.2.5 Musical Agents

Agent comes from the Latin *agere*, meaning "to do" (Russell, 2010). Essentially, anyone/thing that can act with a purpose can be seen as an agent. For example, an agent's sole task might be to recognize the music's particular rhythm while other agents track simple musical patterns, such as repeating pitch intervals and so on (Minsky, 1981). That would be an artificial agent, and since it is concerned with tackling a musical task, it is called a *musical agent*. Tatar & Pasquier (2019) provides an exhaustive literature review of AI and multi-agent systems (MAS) for music. They propose a nine-dimensional musical agent typology. I focus on what they call "Input/Output" (I/O), which defines how the agent listens and outputs within the environment. As the name suggests, it focuses on the auditory modality. I am particularly interested in the input or *perspective* of agents also based on non-auditory interaction channels. Thus, while Tatar & Pasquier (2019) suggest three sub-groups for I/O (symbolic, audio, and hybrid), I make a grouping based on symbolic (e.g., MIDI), audio, affective, and body movement as agents' potential percept features. Although it would have been interesting to include a broad perspective, I will in the following mainly focus on software agents. Embodied agents, on the other hand, are subject to Section 2.3.3.

Symbolic

Chabot et al. (1986) coined the term "composed improvisation" to describe a piece that is shared between non-real-time composition and real-time improvisation. That term also reflects the technological possibilities for a human–machine improvisation system at the time. The system used several algorithmic processes for MIDI-control signals and music analysis (e.g., chords, melodies) but not "intelligence," so to speak. To my knowledge, the first "intelligent" music systems for composition and performance are *M* and *Jam Factory* by Joel Chadabe and David Zicarelli (Zicarelli, 1987). While the former is an interactive composition system, the latter is a real-time improvisation system that listens to MIDI and employs Markov chains as transition tables. Jam Factory consists of four agents—or "players," with the author's words—each holding a pitch set table. The (human) performer has the control of adjusting the probability distribution for the choice of tables. In the playback, another set of tables is used for note durations and algorithms for quantization and rhythmic manipulation (e.g., time distortion, swing).

One of the early examples of musical agents is *Cypher*, which is a real-time interactive music system working in the symbolic (MIDI) domain (Rowe, 1992). It is based on a *listener* and a *player* component. The former comprises multiple agents that analyze and classify the incoming MIDI events based on the register, density, and dynamic. The latter agent can transfer the analyzed MIDI information, generate new material algorithmically, or output a sequence from a corpus of musical events. Here, drawing on Minsky (1986), Rowe (1992) describes each of his hierarchical agent structures as an *agency*, which denotes the unit of agents.

In jazz, there is a performance tradition called *trading fours*, in which one musician improvises for four bars, then another musician takes the lead and improvises for the same amount of time. As most jazz standards have a 32-bar form, four times trading solos accumulate to one full chorus. The "improvisational music companionship" of Thom (2000) is another early example of improvising agents. It focuses on melody generation during solo trading in jazz improvisation. *Band-out-of-a-Box* (*BoB*) operates in two stages. First, it records the training data of the performer offline, possibly during a warm-up session. Then, it uses unsupervised learning based on clustering the histograms of the pitch class, intervals, and melody direction. It builds a probabilistic model of the performer's particular improvisational style. In the second, the real-time stage has "perception" and "generation" components. The former estimates the most likely playing mode for Bob and identifies how surprising the performer's playing mode is. In the follow-up study, Thom (2001) described these components as what makes the system "musically-intelligent."

Audio

A quite well-known musical agent system is the *Voyager* of Lewis (2000). Voyager is a real-time improvisation system consisting of 64 MIDI-controlled agents, operating asynchronously listening and outputting in the audio domain. Lewis, as a notable trombonist improviser and composer himself, points to how Voyager's features, such as its 150 microtonal pitch sets, reflect the particular "multidominance" aesthetic that he inherited from The Association for the Advancement of Creative Musicians (AACM). In technical terms, Voyager stands on the *purely reactive* extreme of the *continuum of autonomy* of Tatar & Pasquier (2019). From the musicking stance, Voyager exhibits a fine balance between moment-to-moment contingencies and a global consistency. The balance lies in the agents collecting small musical details and ideas and reproducing them throughout the performance. The result is that the performer experiences a sufficient familiarity with the structure. Lewis (2000, p. 2) describes Voyager as "a nonhierarchical, improvisational, subject–subject model of discourse, rather than a stimulus/response setup." That reveals a musical autonomy or the opposite of being "purely reactive," albeit strictly programmed rule-based structure. Lewis emphasizes the "de-instrumentalizing the computer" in Voyager, and describes his main motivation as an "anti-authoritarian" impulse. In practice, that resonates well with the bi-directionality of the paradigm shift from open to closed-feedback systems mentioned earlier.

In the *Freely Improvising, Learning and Transforming Evolutionary Recombination* (*FILTER*) system of Nort et al. (2013) the main idea is to build the music system's intelligence on "careful" listening. To that aim, FILTER is built upon the *Deep Listening* practice of (Oliveros, 1984). For example, they focus on two modes of attention: *focal* and *global*. The former points to a critical listening of particular sonic events. The latter is concerned with the entirety of a sound field, a blend of acoustic (e.g., accordion and clarinet) and electronic sounds (e.g., layered delay lines, time stretching, etc.). This

twofold listening principle connects to how FILTER's machine listening is built. Starting from the Gesture and Texture principles of Smalley (1997), FILTER records audio samples of sonic textures and applies a nonlinear time-frequency analysis technique to decompose the signal. Then, using unsupervised learning, FILTER finds textural variations, saves them in memory, and provides feedback about the change. FILTER also incorporates *episodic* and *semantic* memories. Contextually relevant gestures of a longer time duration are stored in the semantic memory. The episodic memory, on the other hand, uses the Factor Oracle (FO) algorithm of another agent-based system called OMax (Dubnov & Assayag, 2005) to learn the temporal structures of gestures. FILTER uses Linear Predictive Coding (LPC), Mel-Frequency Cepstral Coefficients (MFCCs), autocorrelation coefficients, and YIN algorithm for the timbral analysis of (sonic) gestures. It applies a combination of HMM and dynamic time warping for continuous gesture recognition proposed by Bevilacqua et al. (2010). A genetic algorithm (GA) is used to maintain "a globally predictable direction while maintaining random elements on a local scale" by mapping the gesture-likelihood from the recognition process into the GA fitness (Nort et al., 2013, p. 17).

Affective

Among the systems that employ real-time analysis of sound and music features in both symbolic and audio domains, those that use cognitive models, in general, and focus on affective measures, in particular, are relatively scarce. The *OMax* framework mentioned above is a multi-agent human–machine improvisation system that learns from the human performer in real-time (Dubnov & Assayag, 2005). The system proposes a "style injection" in a complex closed-loop manner to balance recurrence and innovation. OMax trains the Factor Oracle (FO) algorithm in real-time based on the audio input of the performers. This algorithm can generate a sequence from the most recent past or jump back in long-term memory. In the authors' words, one of the aims of OMax is to "find a mapping between improvisation parameters and states of the improviser," which they call "mental states" (Dubnov & Assayag, 2005, p. 3). The unique aspect of the system that distinguishes itself from other agent-based improvisation systems is that it builds communication channels between humans and machines in higher-order affective and cognitive descriptors. They use a model that is indirectly related to the mental state of an optimal *flow* experience (Csikszentmihalyi, 1990). Flow, also known as *being in the zone*, happens when a subject is fully absorbed in an activity. In OMax, the *skills* (x) and *challenges* (y) axes of flow is modeled with *familiar* and *emotional force*. In doing so, they can define an optimal experience of surprise, which, even though crucial for improvisation, still requires a moderate amount of familiarity (Borgo, 2002). As such, they also recombined the *Arousal–Boredom* and *Anxiety–Relaxation* states in the original diagram, arguing that the opposite states of arousal and anxiety in improvisation correspond better to relaxation and boredom, respectively. These two dimensions are mapped to the agents' *replication*, *innovation* and *recombination* parameters that control sequences of what is called "factor links."

A more recent example of an improvisation MAS that focuses on the affective aspects is the *Musical Agent based on Self-Organising Maps* (*MASOM*) by Tatar & Pasquier (2017). Similar to FILTER, MASOM also favors and starts from sound-based electroacoustic aesthetics of what can be called free machine improvisation. MASOM's "sound affect estimation" module relies on the implementation of an affect model proposed by Russell (1980). That is a two-dimensional continuous *valence* (x) and *arousal* (y) model, which is trained on a dataset of soundscape samples using multivariate linear regression. Twenty people in an online study made the labeling of the dataset. The affect estimation is based on five audio features, MFCC, loudness, *Spectral Flatness*, *Perceptual Spectral Decrease*, and *Tristimulus* (Tatar & Pasquier, 2017). MASOM uses a similarity matrix to segment the audio material, store it in different files and generate an audio corpus of musical memory. In the following, first, MASOM uses Self-Organizing Maps (SOMs) to cluster the audio corpus. Second, it trains a Variable-order Markov Model (VMM) using a string of SOM nodes. Then, MASOM calculates audio feature statistics using the sound affect estimation mentioned above in the generation phase, and VMM predicts a SOM node to be played next.[14]

Body Movement

The *Robotic Drumming Prosthesis* (RDP) developed by Bretan et al. (2016) is a notable example of shared human–machine control of musical expression and musical human augmentation. The project has been progressed through a design process centered around an amputee drummer's needs, Jason Barnes. The robotic arm evolved through several versions and phases. The first version uses electromyography (EMG) as the primary sensing method. EMG signals are obtained through surface electrodes from two muscle groups (the extensor carpi ulnaris and flexor carpi radialis) of the residual arm. This first version does not use a learning algorithm and mainly relies on real-time onset detection based on a bounded-Q filter-bank decomposition. In the second phase, they incorporated a second stick on the prosthesis as a wearable autonomous agent, which has a "mind of its own," as the authors' remark, and creates rhythmic responses to Jason's and other musicians playing. "A wearable robotic musician extends the notion of a shared interface to that of a shared physical actuator and manipulator," indicate the authors about the new paradigm (Bretan et al., 2016, p. 10). The second stick has two configurations: (1) It can behave similarly to the first stick while the performer controls the initial onset and the subsequent ones by the AI. (2) In the second scenario, AI generates rhythms and timbres harmoniously with the performer's actions.[15] The prosthesis is still developing based on ultrasound technology for sensing. The *Robotic Musicianship Group*

[14] A video footage of a duo performance of the author and MASOM is available at https://vimeo.com/190476284.

[15] A video of Jason Barnes performing with the first versions of the prosthesis is available at: https://www.youtube.com/watch?v=dlSZCu5FAVM

that developed the RDP, have also developed *The Third Drumming Arm* and *The Skywalker Piano Hand* (Weinberg et al., 2020b).

RoboJam is a system that enables a collaborative, improvisatory dialogue that the user can interact with the musical agent using taps, swipes, and swirls on a touchscreen (Martin & Torresen, 2018). The interaction paradigm is a call-and-response improvisation. The user improvises shortly (up to 5s) using the touchscreen interface on a mobile phone. Robojam then generates a response in the server and sends it back to the user's device to play both parts together via different options of synthesized sounds. Usually, the x axis controls the pitch, and y, for effects or timbral features. RoboJam uses a novel mixture-density network (MDN) application combined with a recurrent neural network (RNN), becoming an MDRNN, a generative predictive model. RoboJam's MDRNN integrates the touchpoints on a two-dimensional plane with one-dimensional touch locations indicating the time spent on each touchpoint. That provides the system with a spectrotemporality, which is different, for example, from 2D drawings of *SketchRNN* (Ha & Eck, 2017). The network is trained with a dataset of 20 hours of performance and 4.3 million touch interaction events. After training, a mixture of 2D normal distributions is sampled to generate new locations on the screen, and a mixture of 1D normal distributions, to predict the time. Conceptually, the agent's response is based on the likelihood of the user's next move. However, in doing so, it uses a dataset collected from hundreds of collaborative sessions of different users.

Eingeweide (German for internal organs) is another project that employs muscle sensing and a robotic prosthesis (Donnarumma & Pevere, 2018). It stresses the interaction between the human body and AI from an artistic perspective. The artists focus on an artificial organ that lives outside the human body and is partly independent of human control. Eingeweide features two human performers and a robotic prosthesis placed on the face of one of the performers. The wearable robot uses AI to adapt to the performers' motion while the performer's muscle activity is amplified and transformed into sound. The control-related in-betweenness discussed in Section 2.2.4 emerges with the robot blindfolding the performer such that he cannot rely on his sight on the stage. Instead, he focuses on the auditory and tactile feedback coming from the servo motors of the robot, in addition to other modalities he can use to locate himself on the stage. That sets an example for the co-adaptation of humans, machines, and the physical space. This is an interesting autopoietic configuration, and one that Donnarumma continues to build on in other pieces, such as *Humane Methods* (Vacuo, 2020). In a later article, Caramiaux & Donnarumma (2021) describe the motivation of their work as a conceptual shift from conventional uses of deep learning methods to employing AI algorithm as an actor "performing meaningless calculations." They describe this as (Caramiaux & Donnarumma, 2021, p. 11):

> What interests us is not the capacity of the algorithm to reach its target, but rather the ways in which the inner (obsessive) logic of this type of computation can be made perceivable at an aesthetic and sensorial level to both performers and audience.

Figure 2.12: A press photo of the *Eingeweide* project, with Marco Donnarumma wearing the face-mounted robotic prosthesis. (Donnarumma & Pevere, 2018)

In this context, AI is no longer seen as a tool that improves user experiences objectively. Instead, it starts by questioning AI-powered applications' inherent biases and their social and economic consequences (or drivers). From there, they aim at investigating people's understanding of AI technology. Artistically, they explore the "brute force" of AI algorithms within a choreography that highlights human violence, how it is enforced on human lives and non-human entities.

In Section 2.2.3, I briefly mentioned about the reinforcement learning (RL) system of Visi & Tanaka (2021) for action–sound mappings. A culmination of that system is a collaborative artistic project called *AQAXA* and a music release, *Corporeal EP*, that came on *Punch Up Records* (AQAXA, 2021). The project's main idea is based on sonic memories consisting of audio messages and voice memos recorded over the years. The leading figure of the project and the author of the AI system, Federico Visi, describes his motivation as (AQAXA, 2021):

> Our memories are not linear accumulations of events. What we decide to remember, and what we attempt to forget, shape our personalities. But what happens when we let our machines decide?

The project consists of multiple musicians and several instruments, such as saxophone, percussion, augmented guitar, and various electronics, including a musical agent. As elaborated in Visi & Tanaka (2020), the RL agent, which is

based on *Co-Explorer* of Scurto et al. (2019) (Section 2.2.3), enables a process of exploration of action–sound mappings between the performer's movement and the parameters of a synthesizer that is loaded with audio samples. While exploring the feature space, the agent keeps proposing a new set of mappings and receives feedback from the performer. Ultimately, this iterative process results in a regression model that maps the continuous movement to sound parameters.

2.2.6 Agency

Agency can be defined as the capacity to act in an environment (Russell, 2010; Schlosser, 2019). In discussing the agency of artifacts, Malafouris (2008) formulates the concept using the example of pottery. What is seen from the outside is that the potter's actions give shape to the clay. However, he argues that the causality is two-directional. The energy and motion created by the wheel—the clay is on the wheel—flow back to the potter. According to him, the potter is the active agent here, while the wheel and clay are passive non-agents to be operated and given form (Malafouris, 2008). That is *material agency* and echoes the *gestural agency* mentioned earlier.

Mendoza & Thompson (2017) argue that both a human and a machine can exert over the other within a musical ecosystem. Dahlstedt (2021) stresses a similar perspective by attributing *causal agency* to objects. An object cannot be "blamed" for the outcome of an action executed using them. Starting from there, Dahlstedt defines a complexity spectrum from simple to autonomous tools as depicted in Figure 2.13, based on the *influential agency* that an agent may have.

The above arguments suggest that agents and agency are not necessarily intertwined concepts. Things that are not agents themselves can have or take part of the agency as well. This information will also be handy when discussing the agency of the luthier, composer, programmer, or whoever is not actively present during a music performance.

As first formulated by Hewitt et al. (1973) and further developed by Hewitt (1976), a software agent can be an entity that solely executes specific goals and communicates with other entities according to a script. Although quite simple, that points to an action capability. For example, a thermostat contains a feedback chain in which no human element intervenes (Wiener, 1948, p. 115). When specifying the term agency a bit further in MAS, however, we often encounter that agents are expected to demonstrate some degree of autonomy and intelligent behavior.

Shultz (1991) argues that autonomy denotes movement. First, if an object moves without an external cause, then it is an agent. The movement can be merely goal-directed. Satisfying these goals denotes the ability of the agent to decide how to relate its percept to its output (Maes, 1993). According to Jennings et al. (1998), autonomy also means to operate without the intervention of humans or others. In general terms, we can essentially see autonomy as the opposite of purely reactive stimulus-response behavior, echoing the autonomy continuum of Tatar & Pasquier (2019) illustrated in Figure 2.14. According to Floridi & Sanders (2004), an agent achieves autonomy by possessing two

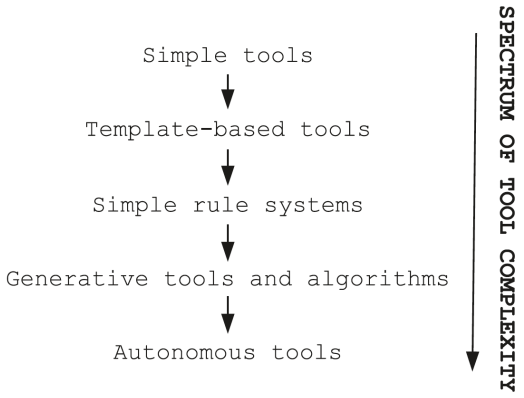


Figure 2.13: The "spectrum of tool complexity" as defined by Dahlstedt (2021). Notice the *tool* on the one extreme end gradually becoming more autonomous towards being what I called *actor*.

Figure 2.14: The "continuum of autonomy" proposed by Tatar & Pasquier (2019); while the systems incorporating strictly pre-defined rules lean towards the left end, the agents that operate with least or non intervention by humans lean towards the other end.

states: respond to others and modify its internal state. That resonates with the notion of *interactivity* that Misselhorn (2015) suggests as the most basic form of intelligent behavior. The intelligent behavior gets more sophisticated as the agent becomes more flexible and adaptive, which, considering complex and dynamic environments, inevitably raises the topic of learning.

Who Makes The Sound?

In the context of embodied music cognition, *actions* are defined as goal-directed cognitive chunks (Godøy, 2018a). Tomasello et al. (2005) distinguishes the external goal from the internal goal, for example, when someone wants to open a box. While the former points to a certain state of the environment as a result (e.g., open box), the latter is linked to the mental representation of the desired state, such as an open box. Possessing internal goals implies what is called an *intentional agency*, which is often considered as a "stronger" notion of agency that covers concepts that are more usually applied to humans, such as intention, belief, or emotion (Wooldridge & Jennings, 1995). The mental representation can also involve higher-order intentions, such as engaging in an activity because you want to accomplish something else. Or, you might have a goal simply because you care about it and find it worth pursuing. According to Helm (2000), that is different from a chess-playing computer of which the behavior is mediated by

instrumental rationality, not by emotions that are intentional feelings concerning personal values.

Even simple automata-like objects can be considered intentional agents (Seel, 1989, p. 80). However, Tuuri et al. (2017) argue that it is necessary to consider the relationship between the embodied mind and the phenomenological environment. According to Gallagher (2007), what distinguishes an agent from purposive behavior is the intentional actions intertwined with a sense of agency. This refers to the feeling of control over actions and their consequences (Moore, 2016). Humans have been developed through natural evolution and, as Legaspi et al. (2019) stress, are the only agents that are fully autonomous. Artificial agents have been created through the intended design of humans, and even if they become sophisticated enough to self-maintain, according to Wan & Braspenning (1996), that is not sufficient to be fully autonomous as they are phylogenetically dependent on their creators.

In philosophical theories of action, not only non-living entities but also infants and animals tend to be considered mere tools excluded from the privilege of the ability to act (Strasser, 2015). In an interactive music system containing both human and artificial agents, the internal and external goals are shared between the author (luthier, composer, performer, etc.) and the agent. Dahlstedt (2021, p. 8) explains this as:

> As art can provoke, I often use the idea of who is the agent behind artistic provocation to sort out what an intentional agent is in art. This acknowledges a sender, an author behind the work, with autonomy and intention.

Human–Machine Collaboration

According to Strasser (2015), sufficient conditions for being an active part of a collective action (often termed as "joint action") is different from those of being an intentional agent. Hence, we may not demand similar abilities of all participating agents. Strasser stresses that we can describe something as goal-directed without intentionality. An action, in the simplest form, requires perceiving information and information processing. This idea was also central in cybernetics right from the beginning (Ashby, 1956). For Strasser (2015), each agent must have a minimal capability of the latter to anticipate the other agent's behavior. Referring to the well-known belief-desire-intention (BDI) model (Rao & Georgeff, 1995), she rearranges the necessary conditions for collective action: (1) The right perception and processing abilities can result in *belief* (knowledge) about how to reach the goal; (2) when the *desire* is initiated, and the goal is transferred to the system, the agent needs to be able to recognize this goal as a goal; (3) a mere goal-directedness of the agent will satisfy the condition of the *intention*. In her formulation of human–machine collective acts, what is crucial is the aspect of high-level communicative abilities. In that regard, she remarks, "[t]o play a role in a collective action one must have *effectors* by which one can

express social hints that are readable for the other agents as well" (Strasser, 2015, p. 12). I will touch upon this aspect more in Section 2.3.

A collaboration between humans and machines (or any agents) can also happen through what is called *emergent coordination*. According to Knoblich et al. (2011), this kind of coordination can occur between individuals without in-advance planning. The phenomenon called *entrainment* (Clayton, 2012) is a clear example of such coordination. Most feedback instruments that noise and experimental music artists use often do not incorporate intelligent agents. The performance, for example, on the no-input mixer mentioned in Section 2.1 heavily relies on emergent coordination. Thus, the performer engages in a "conversation" with the tool, attributing agency to it by synchronizing their actions with the unpredictabilities of self-oscillating circuits. That echoes how Misselhorn (2015, p. 9) describes the interactive behavior of agents as "the behavior of one agent becomes the input of another agent who then modifies its behavior." As a musician himself performing experimental music on electric guitar augmented with motion sensors, Ferguson (2013) stresses the *imagined agency* he attributes to the machine. He describes this as the "invisible and unpredictable presence that acts to stimulate and extend dialogue" (Ferguson, 2013, p. 10). That echoes the "materiality of algorithms" of Goffey (2008), as presented in Dahlstedt (2018). Both approaches emphasize the extent to which one can attribute agency to the other(s) in a particular performance scenario.

Takayama (2012, p. 3) reflects on that "it is still possible to distinguish between what is believed reflectively and what is perceived in-the-moment. In this sense, agency exists in the eye of the beholder." From an aesthetical perspective, then, the question concerning the agency is not only about how autonomous, or intelligent the agent is, nor how much initiative it can take, but also how much it can initiate and what processes it can cause. That echoes the critical stance of the cybernetic vision on arts, particularly the *behaviorist framework* of Ascott (1968). He emphasizes the *process*, what he describes as the dynamic interplay between ordered and random elements. The "feedback loop" principle of cybernetic systems enabled the artistic vision to shift from the field of *objects* to that of *behavior* by blurring the boundaries in the triad artist/artwork/observer (Ascott, 2002). I will discuss that more in Section 5.2.

2.2.7 Summary

A number of authors have carried out reviews of musical AI (Roads, 1980, 1985; Camurri, 1993; Camurri & Leman, 1997; Miranda, 2000; Collins, 2006; Miranda, 2011; Fernandez & Vico, 2013; Fiebrink & Caramiaux, 2016; Tatar & Pasquier, 2019; Miranda, 2021). My aim has been to provide the background to distinguish the use of AI methods *as tool* from their use *as actor*. Since the early 1990s, there has been an ever-increasing trend among artists and researchers working on electroacoustic instruments in employing machine learning as part of the control structures of their musical devices. I provided an overview of the main categories of learning algorithms, tool kits, and example works, and, in doing so,

I categorized them regarding their intended purposes, such as *mapping*, *analysis*, and *generation*.

When it comes to *musical agents*, I categorized some key works in terms of *symbolic*, *audio*, *affective*, and *body movement*. These classes reflect the agents' available channels for perceptual monitoring. That brief review showed that unlike the number of systems using the auditory modality (symbolic or audio), the systems that incorporate motion capture or bio-sensing technologies are highly scarce. The discussion on musical agents eventually led to the topic of *agency*. This concept is understood in many different ways: from minimum requirements, such as *if...then* conditions, to higher-order properties, such as emotions and reasoning. We will get back to this topic from a more embodied embodied perspective in the next section.

2.3 Musical Embodiment

2.3.1 Introduction

Ever since I started using computers for music-making, I have been chasing after an unconventional expression. However, I have missed the *feeling* from my acoustic musicianship. A "feeling" that can give you chills on the stage (Crispin & Gilmore, 2014, p. 131); that can challenge you in playing specific scales (Godøy, 2018b) and thereby motivate practicing; that can facilitate keeping the groove or signaling the drummer to go back to the head of the tune (Jensenius et al., 2010). All in all, this is related to *embodiment*, that is, how the body shapes our experiences (Gibbs, 2005, p. 12), or, more specifically, to *musical embodiment* (Maes, 2016), a notion that most musicians know by heart but rarely think about. As you get more skilled, you process much less information at the cognitive level (Dreyfus, 2001). "Practice your instrument in the air, just by moving your fingers," our jazz ensemble teacher used to tell us, "your muscles will learn."

A dichotomy between body and mind has been prevalent in the field of AI from its inception. As a consequence, one of the favorite investigation areas was natural language (Pfeifer & Bongard, 2006, p. 27). As Thelen (1996, p. 72) stresses, cognitive patterns are dynamic, temporal phenomena that only emerge in the process, not discrete "things" living in the head as symbols and abstractions. It is not the language that gives meaning, Thelen argues: "language taps into prelinguistic meaning." The same situation is mirrored in the music traditions. A similar ontological gap can be found between the symbolic representations of Western music notation and the sub-symbolic sound features that reflect the imprint of human agency (Godøy, 2018b). The combination of both—traditional AI and Western music—yield numerous symbolic musical AI systems that can "compose" in the style of some composers (see Section 2.2.3 for examples). Note that, despite the common use of non-symbolic AI models today, the dichotomy prevails as long as these models are trained on symbolic music data.

This dissertation is grounded on the premise that mental activity and the body are inseparable. We interact with the environment using multiple sensory modalities simultaneously, and cognition recurrently emerges from that

interaction. In this section, I will build on an embodied perspective. First, I will elaborate on *multimodality* and, starting from there, critically portray the evolution of the field of AI. Then, I will clarify the basic terminology of *music-related movement* and introduce some fundamental concepts of embodied music cognition, with a particular focus on *musical interaction*. What will follow is a brief reconsideration of the notion of agency, but this time, from the perspective of the agent, commonly called the *sense of agency*.

2.3.2 Multimodality

Wishart (1996, p. 23) argues that conventional music notation imposes a finite logic upon pitch and tempo, even though these two domains incorporate virtually infinite possibilities. He calls this a two-dimensional lattice. The way the traditional acoustic instruments are built, their mostly discreet nature with keys, holes, and frets, also reinforces that logic. Moreover, the concept of "fixed" instruments adds another dimension of stability to the lattice, making it a three-dimensional one (Figure 2.15).

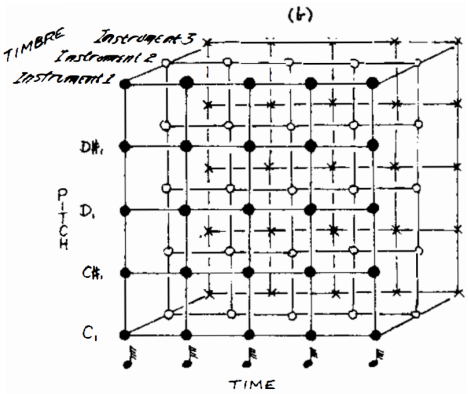


Figure 2.15: A schematic representation of "music on a three-dimensional lattice" as illustrated by Wishart (1996, p. 26)

When playing feedback instruments, such as the no-input mixer (Section 2.1), a distinctive aspect compared to most traditional acoustic instruments is the never-ending contention with the feedback loops. Taming the feedback which can quickly escalate like wildfires (Goldman, 2012), takes more effort than one can imagine due to extreme nonlinearities. That is, a tiny action can cause the system to oscillate close to its breaking point. Feedback musicians develop an embodied knowledge of such practice similar to have acoustic musicians develop their instrumental techniques Tanaka (2015). Moreover, noise music often involves excessive loudness such that it can be a physical threat. It is known that loud music, in general, stimulates the body and causes arousal (Welch & Fremaux, 2017). According to Hegarty (2007, p. iii), the physicality of the loudness, particularly that of the low-frequency sounds, can be favorable

in musical contexts despite the potential of disturbing the digestive systems or heart functions. Donnarumma (2016) shares his musical background in and enthusiasm for the rave culture even though he got his hearing permanently damaged. As a musician, I toured around Europe several times and took part in numerous noise and experimental music events. This has made both familiar with and appreciative of such experiences.

Playing a musical instrument, regardless of the loudness level, involves multiple modalities. These include—in addition to hearing—vision, touch, and mechanoreception (e.g., the relative position of the moving body parts). A striking example of the latter is found in Change Ringing, mentioned in Section 2.2.2, which involves hauling heavy bells in synchrony with other players. Here the proprioceptive mechanism is arguably more important than hearing as it coordinates the moving parts of one's own body in the flow with others. The embodied perspective tells us that we cannot separate the permutative combinations of different bells (pitches), which would be a merely mechanical approach, from the physical coordination and effort in ringing them.

Ecological Approach

Gibson (1979, p. 195), who championed the *ecological psychology*, argues that vision is not a mere channel of sense but a whole perceptual system. He describes how our eyes inhabit our head that is actively re-positioned relative to our posture, yielding the vision. Though Gibson focused on visual perception in the first place, the same goes for all perceptual modalities. Functional magnetic resonance imaging (fMRI) studies have shown that touch can activate a modality-specific area of the visual cortex (McDonald et al., 2001). Our brain can easily be fooled when there is conflicting stimuli, e.g., auditory and visual as in the well-known phenomenon Mcgurk & Macdonald (1976) showed. Consider daily life; even the most straightforward conversation incorporates sound and movement together (Gibbons, 2011), and we understand the speech better if we can observe the speaker's lip motion (King, 1993). We tend to localize moving objects, e.g., a mosquito, by combining visual and auditory modalities as each sense alone is often noisy or unreliable (Churchland, 2011). That echoes Merleau-Ponty (1965, p. 13) indicating, "[w]hen the eye and the ear follow an animal in flight, it is impossible to say 'which started first' in the exchange of stimuli and responses."

Affordance

Merleau-Ponty (1965, p. 13) suggests that our intentions together with the properties of the object "constitute a new whole." This means that we recognize objects based on our embodiment. For example, our motor program of "sitting" is essential to distinguish a chair from a table (Jensenius, 2007). "What they *afford* the observer, after all, depends on their properties," remarks Gibson (1966, p. 285), pointing to our embodied cognitive ability to determine objects' functions based on their shape, size, texture, animation, etc. Gibson's term, *affordance*, is fundamental in the sense of revealing a tight coupling between the

agent and the environment, which defines not only the range of possible actions but also a space for creativity to emerge. This was what Abramović used in her piece where she laid out objects for audience members to use freely during the performance (Section 2.1.3). With a chain, for example, you can make a sound or a weapon.

2.3.3 From Symbols to Body

According to Schank (1984, p. 111), the challenge in getting computers to behave intelligently lies in applying the knowledge the same way we do. Humans, do not linearly perceive, compute, and act. Instead, our perception is coupled with the outer world and continuously adapts, integrating multiple senses. Clark (1999) points to how traditional computers struggle to represent knowledge and reasoning. In the following, I will portray the evolution of AI in tandem with emerging approaches to cognition, from symbolic traditions to embodied and enactive stances.

Early AI

Turing (1950) asked the ambiguous question "can machines think?" To find out whether a machine is intelligent, he proposed a test: the *Turing Test*. Here an interrogator communicates over a teletype, a device resembling a typewriter to send and receive telephonic signals. The objective of the test is for the interrogator to determine whether it is a person or a computer on the other end (Pinar Saygin et al., 2000).

Computers have been used for tasks that require intelligence since the 1940s, the time non-digital computers performed crypto-analysis and trajectory calculation tasks (Brooks, 1991a). Programs back then were based on *brute-force* models,[16] that is, exhaustive coverage of every potential solution to solve a problem. "Less brutish" programs were designed with more elegant search procedures, which can be thought of analogically to solving a crossword puzzle where certain possibilities can be eliminated (Boden, 1977, p. 346). An apt example would be the chess automation developed by Shannon (1950), one of the pioneers of the information theory. Humans do not explicitly consider an exhaustive list of possible solutions to solve a problem. Instead, we base much of our reasoning on the context we are situated in (Dreyfus, 1987).

A notable success came from the Logic Theorist (LT) (recall the "first AI program" mentioned in Section 2.2.2) by realizing the goal that Alan Turing pointed to a decade earlier (Boden, 2006, p. 324). That was to prove theorems from the famous *Principia Mathematica* (Whitehead, 1910). LT proved many of them and even suggested a more elegant version for one. In the sequel of Arthur Samuel's checkers-playing program—beating its creator was already good enough—a program that managed to improve an acclaimed theoretician's work was astonishing. According to Simon (1996, p. 190), LT "solved the mind/body

[16] To my knowledge, *Cyclometer*, invented in 1934 or 1935, is the first device that used brute-force search for decryption purposes (Source: https://en.wikipedia.org/wiki/Cyclometer)

problem." As Simon elaborates, the concept of mind mostly existed only in philosophical discourses. The field of psychology at the time was dominated by the *behaviorist* approach, which, according to Simon, was primarily focused on stimulus–response tasks, largely ignoring the cognitive processes. What Simon and colleagues successfully unrolled was, as the name of their program suggests, logical reasoning. As Boden (2006, p. 924) reports from Moore (1957), the theme *logic* was chosen because Omar Khayyam Moore, who directed the research of the LT, was asking the experimental participants to think loudly in solving their problems. That inspired *protocol analysis*, an empirical research method for studying the cognitive processes of problem solvers (Ericsson & Simon, 1993).

Can a person's reasoning be isolated from the environment and transformed into a data structure? Or, does the foundational capability of reasoning exist in "being and acting in the world" as Popova & Rączaszek-Leonardi (2020) stress? Moreover, aren't these—reason and thought—the things we can know about only through introspection? According to Brooks (1991a), yes, they are. And that is why in the first place he is critical about the *top–down* approaches of traditional AI. In terms of intelligence, he argued that "higher-level intellect" is based on "simple" things in a dynamic environment.

Traditional AI focused on static knowledge structures (Maes, 1993), such as objects and sentences. The sequence of rules that refer to those sentences was constituting the program, which could be stored in memory so that other programs could access as well (Tienson, 1987). This approach succeeded in many applications. However, some tasks that are effortless for humans, such as pattern recognition, are extremely difficult for computers. On the other hand, as Tienson continues, humans are much slower in ample amount of data-handling and number crunching. All in all, he concludes, humans reason differently from conventional computers (Clark, 1999). Thus *good ol' AI* was a different kind of intelligence than that of humans (Tienson, 1987).

Connectionism

Connectionism is a movement in cognitive science that investigates intellectual abilities using simplified models of the brain, such as *artificial neural networks* (ANNs) (Buckner & Garson, 2019). Albeit the dominant symbol manipulation in cognitive science, it was already extensively discussed in the early years of cybernetics that the brain might be based on distributed, massive interconnections instead of rules, central logical processing, or allocated memory locations (Varela et al., 1991, p. 85). As opposed to the traditional information-processing paradigm using symbols and rules, the starting idea of connectionism was units and connections. Each unit or *node* is a parallel computing element, inspired by the neurons in the brain. These nodes are connected to other nodes, thereby inputting and outputting signals from one to the other. We can think of these connections analogous to electric wires, thus having some resistance. When there is less resistance, the signal from one node to the other, hence the association, is more potent (Tienson, 1987). The signal or the information that goes from node to node is presented as *activation values*. The parameters

controlling the numerical strength of each connection are called *weights*, through which the nodes influence the neighboring nodes' activation values. If the weight is negative (*inhibitory*), the influence is negative, and positive, if the weight is positive (*excitatory*) (Smolensky, 1988). In other words, their connection is stronger if both neurons are active; otherwise, the strength of the association is attenuated (Varela et al., 1991, p. 87). This information is then passed throughout the network until it reaches the output. Even though node properties are mostly static in an ANN, the variability of weights provides the network with learning capacity.

The *perceptron* is a single-layer neural network for binary classification (Rosenblatt, 1957; Joseph, 1961; Viglione, 1970). This led to the development of *feedforward* neural networks that use multiple perceptrons, called *multilayer perceptron* (MLP) (Ivakhnenko et al., 1965; Ivachnenko, 1967; Smith, 1980; Hopfield, 1982). This development was also important in music. ANNs brought a shift in algorithmic composition from strict rules to generalizing from learned structures of given musical examples (Todd, 1989). In connectionist systems, finding the right set of weights is crucial to solve the problem. The learning algorithms that can calculate the right weights fall into two broad categories: supervised and unsupervised learning (see Section 2.2.3 for examples). For the latter, well-known algorithms include Hebbian learning, autoencoders, and self-organizing maps (SOMs) (see Section 2.2.5 for different works using these algorithms). For the former, the most popular algorithm is *backpropagation* (BP). Briefly, BP iteratively adjusts the connection weights to minimize the input and output error. According to Schmidhuber (2015), the BP algorithm was first seen in Linnainmaa (1979), but it took another decade to reach its potential by Rumelhart & McClelland (1987).

Today, ANNs, and particularly *deep neural networks* (DNNs) using many hidden layers, dominate AI applications. That dominance, however, did not come until the 2010s. According to Marcus (2018), the image classification DNN implemented in *ImageNet* by Krizhevsky et al. (2012) was the game-changer. Among the critical points about DL, as Marcus (2018) stresses, some deserve to be highlighted. First comes the *generalization* problem, closely linked to the necessity of large datasets. As I will discuss more in Section 5.2, the type of data, hence what is going to be generalized, also poses a problem. For example, the dataset we collected for developing an action–sound model included only one female participant (Erdem et al., 2020). Second, he argues the "shallowness," which he exemplifies with DeepMind's Atari game work that uses DL with reinforcement learning (RL) (Section 2.2.3). According to Marcus (2018), the model masters the game perfectly. However, other than specific contingencies for particular scenarios, it does not learn the physical concepts, such as the tunnel, ball, or the wall. Third, the high accuracy that DL models achieve does not necessarily inhere causality. DL "presumes a largely stable world," Marcus remarks, which echoes several critiques to AI mentioned above.

My position on AI is neither dystopic nor overly optimistic. I agree with the perspective of Clark (1999) that we do not process the world as idealized by symbolic machines. The strategy of building computer models inspired

by the brain seems to have worked. With the emergence of new techniques and faster computers, ANNs not only process information in one direction as in a feedforward network. Memory can emerge through recurrent models. Such models can discover hidden structures in data, find anomalies, synthesize media, inherit social patterns and unroll biases (Johnson, 2020), or define generative processes and behaviors using large datasets (Dahlstedt, 2019, 2021) However, several conceptual and practical issues, particularly the ones that concern embodied interaction and multimodality, remain open.

In sum, classical AI has been successful in tasks such as playing chess-like games, applying rules of logic, and proving mathematical theorems. Connectionist systems, such as DL models, have worked well for image processing, pattern recognition, anomaly detection, and machine translation. Computers still struggle with problems involving embodied interaction, such as talking, dancing, riding a bicycle, or playing a musical instrument. Varela et al. (1991, p. 147–148) argues that these tasks require acquired motor skills and continuous commonsense or background know-how:

> in both cognitivism and connectionism, the unmanageable ambiguity of background commonsense is left largely at the periphery of the inquiry, with the hope that it will somehow eventually be clarified.

Embodied Perspective

The standard cognitivist approach stands with the idea that the brain processes the information that the body's sensory system is equipped with, transforms from one domain, the environment, into a symbolic data structure. Albeit the revolutionary enthusiasm of the embodied approach to cognition, the perspective that the cognitivist holds does not necessarily reject the idea that cognition has its origins in the body and its interactions with the world (Shapiro, 2010, p. 56). Indeed, the problems of isolating the reason and thought from the essentiality of embodied interactions became more obvious in time. That motivated a phenomenological urge for developing concepts and methods to tackle the gap between the world we are situated in and its abstract representations (Dourish, 2001). Thelen et al. (2001) draw a line between the cognitivist and embodied perspectives, emphasizing how the web of "reasoning, memory, emotion, language, and all other aspects of mental life" depends on and comes from a body that has particular perceptual and motor capabilities.

Temporality is also seen as an overarching conception of embodiment (Thelen, 1996), particularly with music's sequential and time-based nature. Clark (2008) focus on the "episodes" during which we experience the incorporation of external equipment. The importance lies in both vocalizing and conceptualizing such interactive processes, which may often involve a temporal "friction," and how the kind of sensory alterations influence the (sense of) embodied agency. As such, the equipment becomes our extensions (Clark, 2004). Moreover, the time delays that often cause those frictions are significant in music interaction with computationally-intensive applications. In CAVI (Paper V), the latency due to

the NN's computation had a significant influence on the music. According to Boden & Edmonds (2009), there is a difference between *interaction*, inheriting a solid action–sound causality, and *influence*, which has rather long-term effects on the output (latency).

The *enactive* perspective shifts the emphasis from seeing the body as an influencer/contributor/partner of the brain to cognition as a unity of brain and body (Varela et al., 1974; Maturana, 1980; Varela et al., 1991). This perspective opposes not only the classic dichotomy between mind and body but also the embodied interpretations that maintains a distinction between the two (Schiavio, 2015). The *enactivist* approach asserts the living body as the cognitive system, regardless if that living body incorporates a nervous system. In other words, the regulation and control of cognition as a homeostatic system are determined by its biological structure (Schiavio & Jaegher, 2017). Hence, cognition is action Varela et al. (1991, p. 172):

> By using the term *embodied* we mean to highlight two points: first, that cognition depends upon kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context. By using the term *action* we mean to emphasize once again that sensory and motor processes, perception and action, are fundamentally inseparable in lived cognition. Indeed, the two are not merely contingently linked in individuals; they have also evolved together.

I find three aspects of this definition important for my work. First, it hints for interactive agent systems to be designed with a sweet spot between *contingency* and *togetherness*. This echoes the perspectives for distributed creativity of Sawyer & DeZutter (2009). Second, the suggestion that the sensorimotor capacities are "embedded" implies that we also share common conceptions of things, such as cultural specifics that concern music. That can explain the constraints in the way some musicians conceive the improvisation. There are some, as Bailey (1993, p. 66) also indicates, for whom the whole activity of improvisation is incomprehensible. Third, it provides a conceptual grounding of thinking a cognitive mechanism without the brain. That is particularly convenient in developing musical agents and subsequently understanding the emerging collective phenomena in the performance, such as the agency.

An example of "embodied AI" (see Kotseruba & Tsotsos (2020) for a review of cognitive architectures) is the robot built by Tani (1998). This robot uses multiple NNs that combine bottom–up sensory–motor processes with top–down predictive modeling of the world originating in the subjective mind. In his approach to AI, Brooks (1991b) advocates a *bottom–up* approach, and one way of doing that was focusing on insect-level intelligence. According to Pfeifer & Bongard (2006, p. 35), it makes more sense to go to human-level intelligence from the insect level when compared to the goal of achieving the intelligence from scratch. As Pfeifer & Bongard (2006, p. 216) elaborates further on the

insect-like behaviors, they point to two important dimensions in the study of collective phenomena: (1) Individual agents can interact in groups that can accomplish things that individuals cannot; (2) global behavior patterns emerge instead of being programmed.

What is called "collective phenomena" of living systems have long been a great interest of cognitive scientists and artists who aim to employ collective emergent dynamics in their artworks. The concepts and methods from the field of Artificial Life (Boden, 1996; Berry & Dahlstedt, 2003; Miranda, 2011; Boden, 2015) have been widely used for musical purposes. Among them, some examples include the system of Martins & Miranda (2007) that uses Genetic Algorithms (GA) for rhythm generation, McAlpine et al. (1999); Miranda (2002) use cellular automata (CA), Dahlstedt (2007) uses evolutionary algorithms (EA) for piano composition, and Beyls (2007) focus on autopoietic self-organization principles for interactive music using GAs.

An early example of using embodied agents in collective multimodal interaction is the emotional agent architecture proposed by Camurri & Coglio (1998). They define *emotional agents* as software agents that possess emotional states, such as anger or sadness. This idea is linked to the idea that emotional content is embedded in the interpretation and expression of the performers' intentions. Agents have three components: Emotional, Reactive, and Rational. In the emotional component, agents perceive different gestures as different emotive stimuli. For example, the catalysts derived from very smooth movements contrast those produced by sharp and nervous movements. The reactive component has little or no state and is primarily employed for various computations. The rational component, as the name suggests, is related to the agent's rational state. That is, agents' knowledge about, for instance, how humans are moving, which virtual instruments have been created, etc. The agents operate in a closed feedback loop. In addition to the audio output, the system also sends messages to the agents for communication. Camurri & Coglio (1998) embodied their architecture in a robot called *The Cicerone* and employed it in a workshop for children. The robot could change its mood according to whatever is happening in the workshop. For example, The Cicerone was getting angry if the games were not played expectedly and expressing itself by changing its movements, voice, music, and environmental lights.

Another relevant example of embodied agents used in artistic contexts are *swarm* robots. Typically, swarms are large groups of insects. In the field of robotics, the concept is used robot collectives inspired by biological swarms (Podevijn et al., 2016). St-Onge et al. (2019) present a swarm system where the robots interact with a dancer in a decentralized manner. The system relies on IMU and EMG signals captured by two Myo armbands worn on the dancer's forearm and calf, capturing spatially more extensive physical states. First, the performer develops three choreographic "moods," varying body postures that reflect different internal (emotional) states. Then, using these pre-defined labels, they collect a custom dataset and train a classifier as the interaction channel with the robots.

A recent musical robot swarm is developed by Krzyżaniak (2021). It

Figure 2.16: A captured moment of interaction between the user and swarm of Dr Squiggles. (Photograph: Kyrre Glette)

consists of three rhythm-playing robots called *Dr. Squiggles*, each of which is equipped with a microcomputer that encapsulates the audio analysis and learning algorithms. Dr. Squiggles receives audio input from other agents. The octopus-looking robot generates rhythms through six solenoids, each attached to a tentacle, interacting with eye motion on a small LED screen. First, we developed an installation piece themed as "air guitar-controlled rhythmic robots," which used Dr. Squiggles controlled by embodied interaction as seen in Figure 2.16 (Section 4.3.1). In the following, a swarm of Dr. Squiggles robots performed with CAVI, the audiovisual agent I developed (see Paper V), and a dancer who interacted through the network. I will mention more about that project in Section 4.3.2.

Weinberg et al. (2020a) provide a good overview of musical robotics. One particularly interesting robot is *Shimon*, an interactive musical robot playing marimba (Hoffman & Weinberg, 2011). It was developed by the Robotic Musicianship Group developed it at Georgia Tech and has performed jazz improvisations for many years. Besides the industry-grade built quality and impressive performance, it also uses *communicative gestures* by head-nodding. In the following, I will focus more on different categories of music-related movement.

2.3.4 From Motion to Gesture

Embodiment in music interaction essentially refers to actions that originate in the body (Leman et al., 2018). As such, the body is the prime medium for interaction. *Gesture* is a commonly used term to describe human motion and has attracted growing attention in music research (Gritten & King, 2006; Godøy & Leman, 2010; Gritten & King, 2011), spanning new musical interactions (Cadoz & Wanderley, 2000; Jensenius et al., 2010; Tanaka, 2011). Set aside the

multifacetedness of the term gesture as reflected in different scholarly fields, it seems that even in the niche of the NIME literature, a commonly held terminology does not exist (Jensenius, 2014). In Paper II published as part of this dissertation, we tried to clarify a basic terminology of gesture. In the following, I will divide the term into different levels of the body movement, for which using a single term—gesture—is confusing.

Low Level

Using a bottom-up approach, I start from low-level body movement, which refers to physical phenomena. Such as *force*, a biomechanical phenomenon capable of altering the state of body motion. According to Newton's second law, force is proportional to *mass* and *acceleration* ($\overrightarrow{F} = m * \overrightarrow{a}$). Force sets the object in *motion*, which refers to the physical displacement of the object. Humans and animals generate voluntary and passive muscular forces to process energy while interacting with the environment (Uliam et al., 2012). Drawing on Newton's third law that two objects at rest exert equal forces on each other, we can relate the push/pull responses to *pressure*. In its equation form, $p = \overrightarrow{F}/A$, A denotes the *area*, an equivalent of which would be the surface of a musical instrument that we mutually exert forces in addition to gravity and friction.

At the moment we touch, for example a guitar fretboard, we immediately *feel* the size of its different parts, what material it is made of, the gauge and wiring of the strings, the height of the bridge, neck, and so on. All these different parts contribute to the transmission of forces, motion, and energy from one to another. For instance, we can approximate the potential energy at the midpoint of a guitar string using the equation $PE = (2Ty^2)/L$, where T is tension, y is displacement and L is length. The experience of playing a musical instrument originates in the sum of the material properties of the instrument and the features of interactive human motion. That is where an expert player is expected to have a complex joint and muscle control (Gonzalez-Sanchez et al., 2019), which results in variations in the energy and frequency spectra of the sound (Schneider, 2018). See, for instance, how the upper harmonics vary by alternating the bow pressure (Motl, 2013), or the amplitude modulation (AM) in *vibrato* effect (Dromey et al., 2009).

Middle Level

The low-level aspects, motion, and force are continuous physical and biomechanical phenomena that can be measured objectively using various sensing technologies (Jensenius, 2018a). However, what has been emphasized in the previous subsection as the (embodied) *action* denotes a psychological experience, a subjective phenomenon. Imagine a guitar player lifting her arm up and then down to hit the strings. Godøy & Leman (2010) refer to "cognitive units" to describe such *chunking* of continuous motion and force. Thus one can think of the action as mental imagery (Godøy, 2009a). As long as an action is not communicated intentionally, it does not necessarily bear a meaning. Hence, I

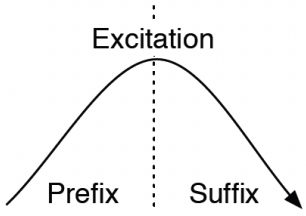


Figure 2.17: An action, such as hitting a guitar string, is realized through an excitation phase, which incorporates a prefix and a suffix. (Jensenius, 2007, p. 24)

place it in the middle level, between low-level physical signals and high-level communicative actions. Since this middle level is subjective, it is impossible to precisely define, for example, the start and endpoints of an action. Consider the case of hitting a guitar string once. As Godøy (2009b) suggests, the attack has an *excitation phase* having a *prefix* (lifting the arm) and *suffix* (lifting down) as illustrated in Figure 2.17. We can define as *fidgeting* the parts of the motion that are not directed by goal nor be intentional or conscious (Figure 2.20).

Considering that both motion and sound are temporal phenomena, we perceive different features in different timescales (Godøy, 2009a). That is a necessity of our cognitive apparatus, for example, in chunking the action segments. Godøy suggests a three-level grouping:

- *Sub-chunk level*: The *micro* timescale for pitch, loudness and timbral features (<0.5 seconds)
- *Chunk level*: The *meso* timescale as well as the timescale for sound-producing actions (0.5–5 seconds) —short-term memory
- *Supra-chunk level*: The *macro* timescale for longer contexts (>5 seconds)— long-term memory

Music-related body motion comes in various types (see Jensenius et al. (2010), for an overview). Here I primarily focus on the *sound-producing* actions. These, based on the typology proposed by Cadoz (1988), can be subdivided into *excitation* actions, such as the right hand that excites the strings on a guitar, and *modification* actions, such as the left hand modifying the pitch. As depicted in Figure 2.18, the excitation action can be divided further into the three main categories proposed by Schaeffer (1966) and presented by Godøy (2006):

- *Impulsive*: A fast attack resulting from a discontinuous energy transfer (e.g., percussion or plucked instruments).
- *Sustained*: A more gradual onset and continuously evolving sound due to a continuous energy transfer (e.g., bowed instruments).
- *Iterative*: Successive attacks resulting from a series of discontinuous energy transfers.

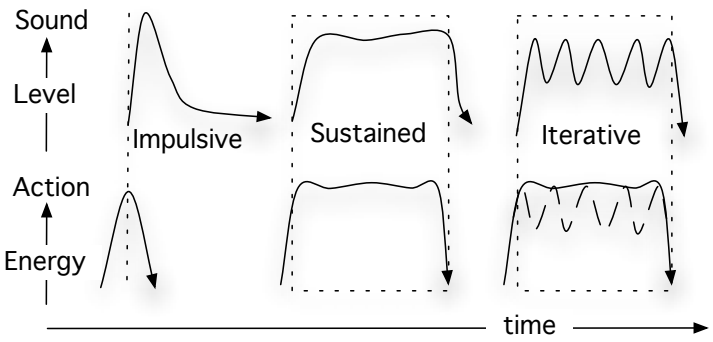


Figure 2.18: An illustration of three categories for the main action and sound energy envelopes resulting from different sound-producing action types. The dotted lines correspond to the duration of contact during the excitation phase. (Jensenius, 2007, p. 26)

Now, once again, consider the guitar player's excitation action for hitting the string. Identifying the excitation phase—investigated in the *chunk* level mentioned above—can be relatively straightforward when dealing with a single impulsive action. However, it becomes highly complex when multiple actions are combined into action series. That leads to *coarticulation* as illustrated in Figure 2.19, the merging of individual actions into larger shapes of actions (Godøy, 2013). That poses a great challenge from an empirical point of view on segmenting, for example, a motion capture recording for motion–sound analysis. On the other hand, in Paper IV, we show that using DL trained with the fundamental sound-producing action shapes (Figure 2.18), we can predict the coarticulated shapes with high accuracy.

High Level

Gestures are actions with an associated *high-level* meaning. The meaning-bearing aspect of gestures has been studied in the field of linguistics: "Gestures exhibit images that cannot always be expressed in speech [...] With these kinds of gestures people unwittingly display their inner thoughts" according to McNeill (1992, p. 12), emphasizing that bodily gestures are essential to the communication.

In the context of music, the term gesture is often used synonymously with both motion and action. However, the challenge is to define the *musical gesture* in a way that covers both motion-related definitions as well as sonic properties, such as the sound-shapes presented by Smalley (1997). Gritten & King (2006, p. xx) does that in a fairly straightforward way:

> [A] gesture is a movement or change in state that becomes marked as significant by an agent. This is to say that for movement or sound to be(come) gesture, it must be taken intentionally by an interpreter, who may or may not be involved in the actual sound production of a

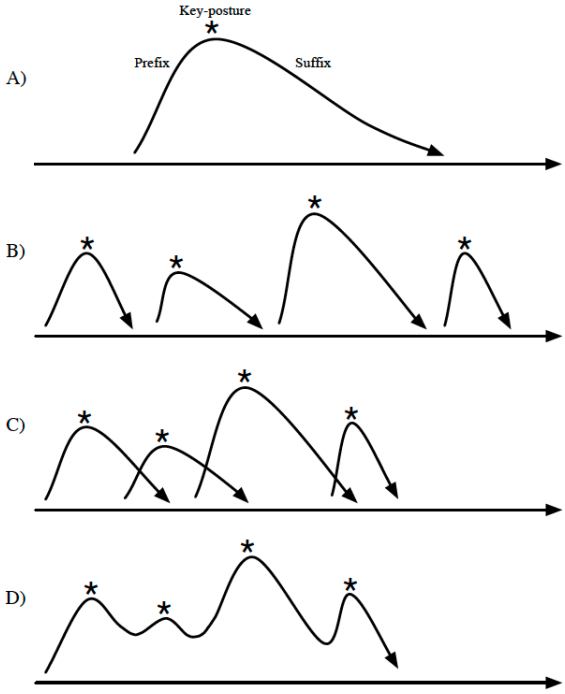


Figure 2.19: A sketch by Godøy (2022) demonstrating how individual action energy envelopes can merge and create new coarticulated actions.

> performance, in such a manner as to donate it with the trappings of human significance.

According to that definition, a musical gesture can be performed unconsciously as long as it communicates meaning to another agent in the environment. That aligns well with the embodied perspective I have introduced earlier in this chapter. The term musical gesture can be related to both motion and sound (the physical), as well as actions and sound objects (the perceptual), as illustrated in Figure 2.20.

2.3.5 Sense of Agency

"AI is not the study of computers," stated Boden (1977, p. xiii), "but of intelligence in thought and action." My attempt at a redefinition would be that "AI is the study of agency." Much of AI research is focused on attributing agency to machines. Thomas Hobbes' poetic words from his famous book *Leviathan* (1651) resonate nicely with such perspective:

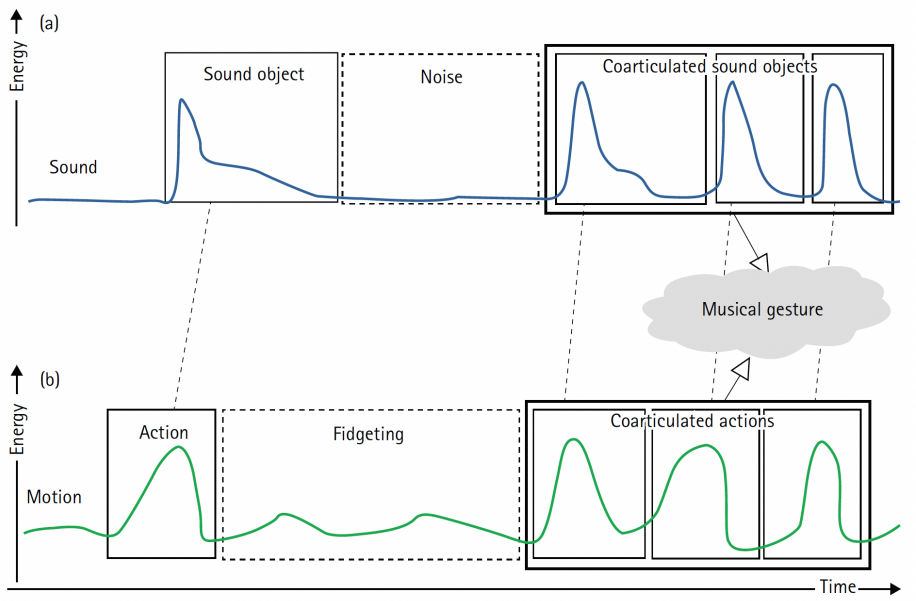


Figure 2.20: A visual summary of how a musical gesture can be thought of as the combination of experienced sound objects (a) and actions (b). These actions and sound objects are perceived from the continuous stream of sound (a) and motion (b). (Paper II)

> For what is the *heart* but a *spring*; and the *nerves* but so many *strings*; and the *joints* but so many *wheels*, giving motion to the whole body, such as was intended by the artificer? *Art* goes yet further, imitating that rational and most excellent work of nature, *man*. (Hobbes, 1929, p. 9)

In Section 2.2.6, I introduced some perspectives on the agency of artificial agents. The minimum requirements for an artifact to be considered as an agent range from a simple *if...then* condition to human-like properties, such as intentions and emotions. But how does one think about human agency? That is a multifaceted topic that has received much interest for centuries. My intention is not to provide an overview of the topic (for that purpose, see, e.g., (Gallagher, 2007; Haggard & Eitam, 2015; Braun et al., 2018)). In the following I will present some concepts to facilitate discussing my work.

The *Self*, first

It all starts from the *self*. Wittgenstein (1969, p. 66) breaks down the "I" into its uses as *object* and *subject*. The first-person pronoun as object can be used in statements such as "I have a bruise on my shoulder," whereas statements, such as

"I think that sounds great," refer to the first-person pronoun as subject. According to Wittgenstein, the former is prone to errors, e.g., illusions or manipulated perception. One can reasonably ask, "are you sure that it is a bruise?" However, in the latter, even though the person might be wrong or even delusional about the statement's content, she cannot misidentify herself as the "I" who state that. In other words, it does not make much sense to ask: "are you sure it is you who think that?" Shoemaker (1984, p. 8) describes the first-person statements as subject "immune to error," whereas "I" as object is not, because in the former case, as Gallagher (2000) also emphasizes, the access to the self is non-observational.

Gallagher (2007) elaborates on the phenomenology of the sense of agency at two levels: (1) The *sense of ownership* (SoO), which is the knowledge of "this is my arms and legs that are moving;" and the *sense of agency* (SoA), the sense of "it is me who initiate or cause the action." That is necessary in the case of involuntary or unconscious actions. As Gallagher maintains, one can think: "I am confident that I own my movement since I know that my body is moving, whereas I may not feel that is me who cause or control the movement, thus not feel agency." For example, if a vocalist approaches a loudspeaker too closely with a microphone, it can cause acoustic feedback. The vocalist then would own the situation in which she is holding and moving with the microphone but not controlling the consequences of the acoustic feedback. Hence, there is a difference between one's authorship of the effects of her actions (SoA) and one's ownership of intentions and (body) movements (SoO) (Gallagher, 2000, 2007; Sato & Yasuda, 2005; Braun et al., 2018).

Alvin Lucier's "brain wave" piece, Music for Solo Performer (see Section 2.1.3), is a relevant case for this discussion. He is the composer; therefore, he owns at least the broad plan of the musical scenario. He is also the performer, hence owns whatever change he brings to the world's current state. However, the movement of the percussion instruments based on his amplified brain waves is mainly out of his control. Therefore, according to Gallagher's distinction, Lucier's SoO should be high while the SoA is low. Artistic contexts, especially those that incorporate machines with unconventional control structures, can be highly obscure for the SoA. So how do we determine the agency?

Sensing the Self

A prominent theory that addresses agency is *apparent mental causation* (AMC) (Wegner & Wheatley, 1999; Wegner, 2002, 2003). It suggests three principles:

- *Priority*, that is, whether the thought of the action precedes the corresponding action or not. A somewhat equivalent would be the timing in music.

- *Consistency* between the intention and the action. In music, that aligns well with the causality in action–sound relationships.

- *Exclusivity* of the source of action, or, in other words, whether there are other sources that can potentially create ambiguity. We can think of this more in terms of musical control.

In his theory, Wegner argues that just the thought of an action can be enough for the experience of agency. That is the case even if the action is not performed by that person. In other words, if we just had a conscious thought about an action, which turned to be consistent with the incident after it occurs, that can create an illusion of us being the cause of the incident. The emphasis in such a mind trick is put on the perception of causality, for which consistency is fundamental. Wegner (2002, p. 79) remarks:

> The principle of consistency operates in apparent mental causation because the thoughts that serve as potential causes of actions typically have meaningful associations with the actions.

van der Wel et al. (2012b) indicate that Wegner's theory can fall short in explaining the interaction between perception and action. Instead, they stress the importance of both sensorimotor and perceptual cues that can determine the agency. In the former, the SoA is related to an automatic internal comparator mechanism called *forward* prediction model (Wolpert et al., 1995). When we perform an action, the model uses the signal generated by the motor system (*efference copy*) to make predictions about the outcome. In turn, the model checks the congruency of incoming (*reafferent*) sensory signals (e.g., visual, proprioceptive). If the results match, then we sense agency, a feeling of having authored the action (Haggard, 2005; Jeannerod, 2008; Gentsch & Schutz-Bosbach, 2015). For example, when you grab your guitar, you sense agency; when you hit the string, you feel the same. However, playing your guitar through the sound card on a computer with perceivable latency, your SoA will be lower due to a violated response time expectation.

Several studies that investigated the discrepancies between actions and visual feedback reported that in the case of distortion in the feedback (Farrer & Frith, 2002) or angular bias (Farrer et al., 2003; Synofzik et al., 2006), participants tend to attribute agency to others. In Paper VI, we report on an online study in which participants attributed more agency to a visual widget that incorporated angular bias and drifting Brownian motion compared to other animated conditions with more direct control. Haggard et al. (2002) reported what they called an *intentional binding* effect. This is based on perceiving actions shifted forward in time while perceiving the outcome moved backward. That study provided additional proof to the neural correlates of the importance of temporal (in)congruities between action and its consequence.

How can we think about SoA in joint activities, such as collaborative improvisation co-performance? Georgieff & Jeannerod (1998) reported overlapping activations of regions of the cerebral cortex during subjects producing actions and observing action production. Similarly, Fourneret & Jeannerod (1998) compared two conditions of visual feedback, in which subjects demonstrated a

tendency of following the bias introduced by the computer instead of sensorimotor cues derived from their own movements. According to van der Wel et al. (2012b), both studies confirm the importance of perceptual cues. From there, they stress the potential confusion in collaborative scenarios, "the perceptual consequence of an action may stem from one's own or another's actions." Jeannerod & Pacherie (2004) describes the issue as a "self-recognition problem," and both the self as subject and object can experience that even though the former has been widely accepted immune to errors. He concludes by suggesting two levels of self-recognition: (1) A "subpersonal" (automatic) level that immediately demonstrates adaptive abilities; and (2) "personal" (conscious) level that represents intentions, plans, and desires of the agent. In that regard, Jeannerod & Pacherie (2004) argue:

> The question here is to determine what are the cues a subject uses to build his conscious sense of being the author of his own actions (the sense of agency); and, more specifically, to determine to which extent the automatic mechanism can contribute to this sense of agency.

van der Wel et al. (2012b), however, disagree with Jeannerod & Pacherie (2004), arguing that perceptual cues are mostly processed automatically. In a follow-up study, van der Wel et al. (2012a) examine the SoA over dyadic joint actions. There, participants tried to bring a pendulum to equilibrium by pulling on chords attached on each side, while the experimenters measured the participants' exerted forces. Following the comparison of the data from force sensors and the participants' feedback regarding the performance quality, the results showed that agency judgments relied most strongly on the perceived quality of the shared performance and not on sensorimotor information (predictive incongruity). These results provide additional support for the emphasis on the close relationship between perception and action, or, on the "perceptually-guided action" (Varela et al., 1991, p. 173).

Perception–Action Loops

Several of the embodied theories put a strong emphasis on the coupling of perception and action. In a nutshell, we can read that perspective as a circular organization in which our actions (goals) influence the contents we perceive (feedback), affecting the actions we take, and that goes on and on as a continuous feedback loop as in the control-systems principles (e.g., see Wiener (1948); Ashby (1956)). The discovery of *mirror neurons* in the macaque brain showed that both executing and observing an action leads to the same neuronal activity, which signaled a fundamental shift. Following the works of Gallese et al. (1996); Liberman & Mattingly (1985), we know that the same areas of the human neural system activate in both performing an action and perceiving someone else doing the same action (Rizzolatti & Arbib, 1998; Hickok et al., 2003).

In the *common coding theory* (CCT), van der Wel et al. (2012b) suggests that both action and perception use a common representation (code) in the brain.

For example, if one of two persons made an action while the other watched, they would both have similar action representations. The studies they report show that, for example, expert athletes outperform non-player watchers in predicting the outcome of a basket shot before the ball leaves the hand. In other words, players are better at reading body kinematics and then making predictions about the result of the action. That resonates with studies investigating relationships between auditory and motor perception (Godøy, 2003; Godøy et al., 2006; Lahav et al., 2007).

Adding Sound-Making in The Loop

The perception of causality is crucial in music performance and perception. The sound production on a traditional instrument is bound by the physical constraints of the instrument and the capabilities of the human body. The physical properties of an instrument define its unique timbre and playfulness. Godøy (2018c) argues that the human body has certain biomechanical limitations that are part of the transformation of embodied action into sound features. Jensenius (2007, p. 23) defines these transformations as *action–sound couplings*, the relationships that abide the laws of physics.

In contrast, electroacoustic musical instruments are based on the creation of *action–sound mappings*. Here the hardware or software constraints are often open to interpretation. In other words, the relationships between biomechanical input and the resultant sound are designed and may not correspond to each other. Echoing the consistency principle mentioned above, however, the creation of meaningful action–sound mappings is critical for how an instrument is played (action) and how it sounds (perception) (Hunt & Wanderley, 2002; Van Nort et al., 2014). That is often discussed as the "mapping problem" (Maes et al., 2010), which has been a central research topic in the field of new interfaces for musical expression over the last decades (Jensenius & Lyons, 2017).

Landing The Self's *Exclusivity*

Exclusivity is the last principle in Wegner's AMC theory. He accounts some puzzling cases to exemplify the violation of that principle. It is common for patients suffering from delusional states or schizophrenia to misattribute their wills to an external agency or force (Sato & Yasuda, 2005). According to Wegner & Wheatley (1999), it is also common among healthy people who dowse for water to report that the forked stick moves by itself and not by the will of the person holding the stick. In a study they refer to, Vogt (1959) observed people experiencing the loss of voluntariness due to the unpredictable movement of the dowsing rod. In a completely different context, the subjects identified their hands as "anarchic" upon observing delayed visual feedback in a study by Leube et al. (2003).

In his book *The Illusion of Conscious Will*, Wegner (2002, p. 99) exemplifies some more extreme cases. An interesting one is the Ouija Board, which usually comes in a planchette and looks like a regular board game. According to the

instructions, two or more users concentrate on a question as they have their fingers on the planchette, waiting for an unintended movement, often from a supernatural being.[17] The whimsicality aside, Wegner (2002, p. 110) stresses the *sense of involuntariness* within such a social magnification of automatism. He points to two possible drives behind that: One option is the unpredictability introduced by the presence of others. Co-actions can reduce one's perceived causality between her thoughts and the observed movements. The movement trajectories introduced by another person may not be consistent with one's initial thought. The other option is that being in a group can result in less conscious intentions, thus making the individual lend her own intentions. Wegner (2002, p. 113) remarks:

> So the action may seem to arise without prior thought. It may even be that when a nonself source is in view, this neglect of preview thoughts leads the individual to become less inclined to monitor whether the action indeed implements his own conscious thoughts.

That is an exciting take on agency, or the lack thereof, considering some musical examples I introduced in previous sections. For instance, we can read the coadaptive performance approach of Tanaka & Donnarumma (2018) as if the performer is *landing the exclusivity*. Can that be a deliberate (non)control approach? While the performance pieces in Section 2.1.3 are some obvious cases, we can see similar urges in a variety of collaborative performance practices, e.g., free jazz and improvisation (see Section 2.1). With that in mind, one can find meaningful parallels with new interactions that can emerge during collaborative performances.

Emergent Coordination

Collaborative activities of two or more agents lead to *joint actions* and often denote a joint or *shared* intentionality. Or, said, differently, multiple people coordinate their actions to bring change to the world's current state (Sebanz et al., 2006). Often in joint actions, it may not be possible to distinguish individual actions (Woodworth, 1939). According to Bratman (1992), a joint action requires mutual responsiveness, a commitment to the joint activity, and a commitment to support each other. Tomasello et al. (2005) investigate such actions in terms of shared intentions, with a particular focus on infant behavior from an ontogenetic perspective. Ecological psychologists focus on the dynamics (Marsh et al., 2009), and some cognitive psychologists, on the embodied aspects (Sebanz et al., 2006).

Among other proposed coordination models, Knoblich et al. (2011) distinguish *planned* and *emergent* coordinations that can occur during joint action. In the former, as the name suggests, the representation of the outcome is essential. As Tomasello et al. (2005) suggest, planning, and being persistent about

[17] Wegner (2002, p. 221) discusses that in terms of people's action projection to imaginary agents, which he describes as *virtual agency*.

the plan are among fundamental aspects of intentional actions. Infants develop an understanding of these concepts from quite an early age. In the emergent coordination, however, coordinated behavior arises without any prior plan. Knoblich et al. (2011) suggest four situations that can source emergent coordination (EC):

- *Entrainment*: The physical phenomenon of the synchronization of multiple independent rhythmic processes (Clayton, 2012; Leman, 2012), such as people clapping together in synchrony (Néda et al., 2000).
- *Common affordances*: As elaborated in Section 2.3.2, affordance denotes the action possibilities of an object (Gibson, 1979). Then, common affordances represent situations in which two agents of similar action repertoires encounter objects that afford similar possibilities, e.g., two percussionists and a drum set.
- *Perception–Action Matching*: Similarly, this one also includes multiple agents with similar action repertoires. A matching emerges when one observes the other one's actions as being familiar. For example, a percussionist might start playing "air drums" if someone else is doing the same thing.
- *Action simulation*: Recall the example of the expert basketball player mentitoned above; observing from outside, she can predict the outcome of a shot better than the expert fans as soon as the ball is about to leave the player's hands. According to Knoblich et al. (2011), that predictive ability can lead to emergent coordination as similar expectations can induce similar action tendencies.

Collaborative improvisation almost always enacts these situations. After all, the concept of emergence is crucial in improvisation (Bailey, 1993; Borgo, 2005; Kosowitz & Vickery, 2013). In a study by Hart et al. (2014), dyads were given a task based on the "mirror game" (Noy et al., 2011), in which the participants improvised by moving parallel sliders to create expressive patterns. The results showed that smooth, synchronized, and complex motion emerged during the performance of expert improvisers. A follow-up study demonstrated how such a joint flow, which the authors describe as "togetherness," led to an increased correlation of players' heart rates and increased motion intensity (Noy et al., 2015). Recalling the study by van der Wel et al. (2012a) where shareable perceptual cues were shown strongly efficient, such aspects of collaborative practices must have a pivotal role in the experience of the agency. One can then argue that if the mind can extend (Clark, 2004), so can the sense of agency (SoA). As such, Moore (2016) points to the flexibility of SoA, indicating that the cases in which we experience agency can be quite incongruent with the facts of the agency. He exemplifies such extreme cases with voodoo dolls and how people can genuinely believe that they can cause harm to some other people by sticking needles remotely. He suggests (Moore, 2016, p. 2):

> So the flexibility that might make us vulnerable to agency errors in things like placebo buttons and voodoo dolls can also allow our experience of agency to extend into new domains and track the rapidly changing agentic structure of our environment. Rather than our agency processing system breaking down with the development of tools, which have changed and extended our agentic capabilities, it has been flexible and adaptable, allowing us to accommodate these changes.

2.3.6 Summary

In this section, I reviewed some significant developments and directions in AI and cognitive sciences through the lens of musical embodiment. The *musical* part of it did not necessarily aim at focusing on music but using it as a basis, or an attitude, a starting point to understand the trends in technology and science. Following discussions of multimodality and affordances, I compared *symbolic* and *connectionist* traditions. My aim was to make the embodied perspective apparent. Interestingly, the connectionist approach has grasped some essence of how the human mind works, producing fascinating scientific and artistic tools. However, I doubt if it will ever be near clarifying commonsense ambiguity, nor being creative itself, through prevalent monomodal approaches. With that in mind, I went into the *embodied perspective*, where I introduced a few varying stances/interpretations of embodiment, thereby finalizing the focus on conceptions akin to AI.

In the following, I clarified terms including *motion* and *force*, both physical phenomena; *action*, which denotes goals, thereby pointing a psychological experience; and, on top of all, gesture, which bears *meaning*, a communicative component with respect to a higher-level consciousness. Such different aspects of body movement, which I leveled as *low*, *middle*, and *high*, respectively, are crucial for two reasons: (1) In developing the perceptual monitoring systems of artificial agents, and (2) in understanding the concept of agency for both human and non-human entities.

Finally, I discussed the *sense of agency* and the notion of *self*. Central topics here include *causality*, *sensorimotor* information, and *perceptual* cues. All in all, the common emphasis was on the *perception–action*, which was subsequently connected to *action–sound relationships*. Then, I questioned the *exclusivity* of individual (musical) agents and stressed the importance of *emergent coordination* in joint activities. These terms and perspectives are pivotal as they incorporate close links to kind of artistic and musical vision I have referred to since the beginning. In the next chapter, I will present my methodology, framed by the concepts and theories introduced here.

Chapter 3

Methods

Wonderful things would come out of that box
if only we knew how to evoke them.
– *J.R. Pierce* (Pierce, 1965)

3.1 Introduction

As described in the previous chapter, this dissertation builds on many different theoretical perspectives. Over the last years, I have also employed a number of methodologies: literature review, observation studies, experiments, design, development, prototyping, performance, and evaluation. My overarching approach can perhaps best be described as *iterative prototyping*. I have continuously moved between creative and reflective modes of working. My research has been structured around four projects, each of which resulted in an interactive music performance framework:

1. Vrengt, an interactive dance piece in which two performers share the control of the system (Section 3.2)
2. RAW, a muscle-based instrument exploring a chaotic behavior in control, and automatized ensemble interaction (Section 3.3)
3. Playing in the "air," a predictive action–sound model using deep learning based on a custom dataset collected throughout a series of laboratory experiments (Section 3.4)
4. CAVI, an agent-based interactive system using a generative model trained on the data collected in the previous study (Section 3.5)

In the following sections, I will elaborate on the methods used in each project; my intentions, the logic behind the developed systems, and their outcomes. Considering the number and variety of methods employed and tools used, I have grouped them under specific categories as depicted in Table 3.1. At the end of the chapter, I will reflect on my methodology more broadly.

		PROJECTS			
		Vrengt	**RAW**	**Air Guitar**	**CAVI**
INTERACTIVE SYSTEMS	**Sensors**	EMG IMU Breathing	EMG IMU Audio	EMG	EMG IMU Audio
	Control	Fixed Collaborative	ML Rule-based Chaotic	ML Predictive	ML Rule-based Generative
	Generation	Sonification Live EFX	Sonification Live Sampling Live EFX	Sonification	Live EFX Sonification Visuals
	Prototyping	Python Max	Python Max	Python Max	Python Max/Jitter
	Performance	Comprovisation	Improvisation		Improvisation
EVALUATION	**Data Collection**	Interview Audio Video Self-report	Audio Video Self-report	EMG IMU MoCap Audio Video	Interview Questionnaire Audio Video EMG IMU Breathing
	Analysis	Qualitative Subjective	Subjective	Statistical	Qualitative Observational

Table 3.1: An overview of the methods and tools used in the dissertation, organized according to the four included projects. Notice the top and bottom panels, which group the methods employed for *interactive systems* and *evaluation*, respectively. Abbreviations: Electromyogram (EMG), Inertial Measurement unit (IMU), Effects (EFX), Machine learning (ML), Motion Capture (MoCap).

Figure 3.1: A collage of captured moments from the rehearsals. Notice that the system allowed collaboration with others. We did so with a visual artist, a rehearsal with whom can be seen on the left; and with a second musician, who is seen in the top-right picture.

3.2 Vrengt

Vrengt (the Norwegian word for "inverted") is an interactive system that allows a dancer and a musician to share the control.[1] The shared aspect required a different approach than standard sonic interaction design. That is to make the design process as collaborative as possible so that the musician and dancer do not work in separate "layers." First, the development was made using a participatory design approach. That was a highly integrated process of fast prototyping, trials, rehearsals, data collection, analyses, re-design, conversations, recording sessions, and subjective evaluations (Figure 3.1).

3.2.1 Design

The design of *Vrengt* was based on a circular organization: capturing and sonifying the dancer's (micro)motion and the shared control of the sonification parameters, which in turn affected the dancer's motion (illustrated in Figure 3.2). Human *micromotion* can be seen as the tiniest producible and observable motion (Jensenius et al., 2017). Numerous physiological and biological processes that we execute unconsciously for executing actions are often manifested as micromotion (Chi et al., 2000). The idea was to work on *sonic microinteraction*, an interaction mode that is common in acoustic instruments, but rarely found in interactive systems (Jensenius, 2017).

[1] A video teaser is available at https://youtu.be/vXJ0l9Q68nc

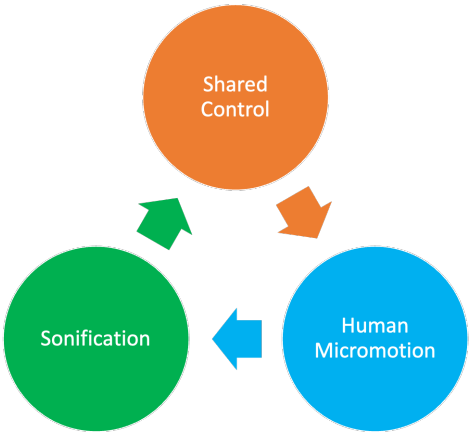


Figure 3.2: A diagram illustrating our design perspective and the conceptual flow of the system as the basis for the implementation.

In *Vrengt*, we used muscle sensing through electromyograms (EMG). The EMG signal captures human micromotion indirectly, since this level of interaction not always result in overt body movements (Tanaka, 2015; Jensenius et al., 2017). EMG is often able to report small or non-visible motion akin to both consciously executed actions and automatic processes of the body (Ortiz et al., 2011). As for the specific sensor device, we chose to work with the (at the time) commercially available *Myo armband*. This is a reliable and cheap solution compared to medical-grade devices (such as *Delsys Trigno*) that I used for lab-based experiments. The benefit of working with Myo was that we could experiment by hacking it to adjust to the calf of the dancer (Figure 3.3) as we wanted two armbands to be worn on both upper and lower body parts for a better whole-body experience. Second, it is a plug-and-play device with easy-to-use software interfaces developed specifically for interactive purposes (see Section 2.1.4). It was particularly convenient to build on the *myo-to-osc* interface developed by Martin et al. (2018b). Here we upgraded the Python scripts (Lan, 2019) to support individual Bluetooth Low Energy (BLE) adapters to overcome possible bandwidth limitations. Fortunately, since Python is one of the most popular programming languages these days, it comes with a huge community and resources for this type of development.

The second interaction method employed in *Vrengt* was capturing the breathing of the dancer in the form of audio signals sent through a wireless transmitter. The reason for choosing this body signal and capturing method was twofold. First, breathing also reflects an in-betweenness between bodily control and automaton similar to muscle activations. Second, we preferred using audio over a dedicated respiration sensor so that the dancer could use the headset microphone to deliberately create acoustic feedback loops by changing her proximity to the speakers on the stage. In doing so, the breathing was also used as an aesthetic element on its own. Since the dancer's position on stage influenced the produced sound, the physicality of the space became an integral

Figure 3.3: The modified Myo armband to fit around the calf of the dancer.

part to the performance. This was particularly effective in the opening of the piece, when the dancer was blindfolded. Then she had to rely on the auditory feedback from the system to orient herself.

3.2.2 Sound

Sonification was a core method used in the sound design of *Vrengt*. This approach was chosen to give the dancer a direct and immediate sonic response. Sonification is often seen as an objective approach to representing data through sound (Hermann & Hunt, 2011). In our context, sonification was not the end goal. Instead, we used sonification as part of the creative process.

We intentionally focused on two techniques in the sound design: (1) Physics-based synthesis of everyday sounds and (2) abstract techniques. One of the physic-based synthesis models was a radially oscillating bubble model, which can be represented as:

$$l(t) = a\sin(2\pi ft)e^{-dt}$$

$$f = 3/r$$

$$E = 2pr^3u^2,$$

where the physical parameters of the bubble radius (r), damping (d), the amplitude (a), the liquid density (p), time (t) and velocity (u) of average inward motion are available and can easily respond to physical input parameters (Doel, 2005). In doing so, we could also explore the dancer's perceived sensations concerning the sonic imagery of the particular sound synthesis techniques and the mappings.

As for abstract techniques, we explored waveshape distortion, ring modulation (RM), and exponential frequency modulation (FM) in the *Sound Design Toolkit* (SDT) for physically-informed procedural sound synthesis (Baldan et al., 2017). SDT was a recent package that mainly focused on everyday sounds and interactions (e.g., friction, liquid sounds, etc.). According to the dancer, while physics-based sounds evoked a more straightforward imagery, the use of abstract techniques for sound synthesis resembled shapes that she could "fill with any image you want."

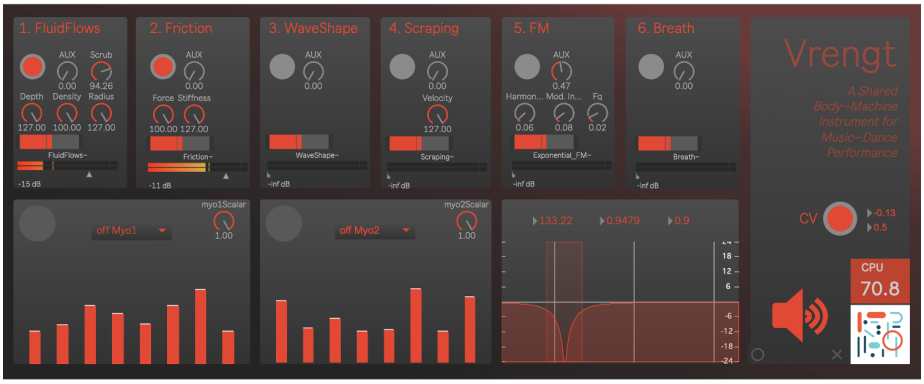


Figure 3.4: The graphical user interface (GUI) of Vrengt, designed in Max.

3.2.3 Interface

The software part of *Vrengt* was developed in *Max/MSP/Jitter*, a graphical programming environment for data, sound, and visuals (Puckette, 1985; Zicarelli, 1998). I considered developing in Pure Data (Pd) instead, another graphical audio programming environment also developed by Puckette (1996). Pd is free, open-source and has many available objects and models. However, for this project we found that Max was beneficial due to its extensive documentation, external libraries, maintenance, tutorials, and flexibility when making graphical user interfaces (see screenshort of the GUI of Vrengt in Figure 3.4).

3.2.4 Mappings

When it came to mappings, we decided to work with fixed mappings in Vrengt. This was decided early on to accommodate that two performers would share the control. The dancer's incoming sensor and audio data were processed and interpreted in real-time by the musician, who used knobs and faders on a MIDI controller. This way, both performers could experience the other's agency. This was perceived as inspiring by both performers, and fuelled further implementation of artificial agents. However, while the dancer was interacting using multiple modalities, the interactive channels of the musician were rather scarce. Reflections on the experience of each performer were collected and presented in Paper I. The performance was what can be called a *comprovisation* (Dudas, 2010), an improvisation that is directed via pre-composed elements.

3.3 RAW

In Section 2.1.4, I introduced how the biofeedback paradigm was used by experimental musicians from the 1960s and onwards, moving to biocontrol in the 1990s and then into coadaptation in the 2010s. In designing *RAW*, I prioritized the latter, that is, focusing on an interaction paradigm where the

system not only adapts to the performer but the performer is also expected to adapt to the system and its physical environment.

The collaborative improvisation of *RAW* built on the shared control of Vrengt but with a focus on co-adaptation. Imagine a double pendulum and how small changes in the initial angle, mass, and speed conditions can influence the overall motion. In an improvisation ensemble, the performer's action is often strongly influenced by other agents within the environment, whether a human performer, an audience member, or a machine. In other words, intentional actions can be enriched or challenged by energy influxes and moment-to-moment contingencies. In RAW, I aimed to simulate that aspect within the control structures of an electroacoustic instrument.

3.3.1 Muscle Sounds

The name of *RAW* comes from the system's primary distinctive property: It uses raw bioelectric muscle signals (EMG) at audio rate (Paper III).[2] This was inspired by *Myogram* by Tanaka (Tanaka & Donnarumma, 2018), using a direct audification approach showcasing the performer's continuous visceral activity. In the performance setup, two Myo armbands are worn, one on each forearm. Four EMG channels (two per forearm) are buffered at every quarter of a second, which has been found to be an acceptable latency threshold (Englehart & Hudgins, 2003). The buffered EMG is then converted to an audible level by increasing the frequency via a time-scaled sawtooth signal. In doing so, the inherent noise of the raw signal is also frequency-shifted, thus creating a quite noisy high-frequency layer in the audible spectra, requiring filtering. This phase is where the the performer can start being creative as a composer. For example, speeding up the signal to extreme values introduces glitches reminding of well-known electronic music textures, such as those of Ryoji Ikeda.

Two channels of EMG per forearm are sonified, corresponding to extensor and flexor muscle groups. This provides four channels of drone sounds, which are controlled by the extension and flexion of each wrist. Other poses, such as ulnar or radial deviation, open or closed hands, and neutral poses, create different combinations. One can imagine such a scenario as mixing four audio channels using faders on a mixing board. This simple approach can be awe-inspiring when used with an extensive, multi-channel sound system. However, it can also fall short of more sustainable use in different ensemble settings, which was one of the main drives of the project in the first place. To that aim, I explored several algorithmic approaches for generating control signals.

3.3.2 Control

In the control part of the system, I used multiple feature extractors simultaneously. First, amplitude envelopes were extracted as the root mean square (RMS) of the continuous EMG signal. The moving RMS of a discrete signal is defined by

[2] A video teaser is available at https://youtu.be/_--dzA5pl9k

St-Amant et al. (1996) as:

$$\hat{\mathrm{x}}_1(\mathrm{t}) = \left[\frac{1}{\mathrm{N}} \sum_{\mathrm{i=t-N+1}}^{\mathrm{t}} \mathrm{m}^2(\mathrm{i})\right]^{1/2} \tag{3.1}$$

where $\hat{x}$ is the EMG amplitude estimate at sample t, using a smoothing window length of N. That works well for larger-scale events. However, for time-sensitive operations, such as triggering, I used the Teager-Kaiser Energy (TKE) operation to calculate the muscle onsets, which is defined in the time domain by Li et al. (2007) as:

$$\mathrm{y(n)} = x^2(n) - x(n-1)x(n+1) \tag{3.2}$$

The RMS is a simple feature that can efficiently be used for the control of dynamics. For goals requiring more precision, such as promptly triggering an event, I prefer to use the IMU data, particularly the *jerk*, the rate of change of the acceleration. However, not for precise control purposes but to create pointillistic percussive sounds within the texture, muscle onsets calculated with the TKE operation can be used. Furthermore, I trained a support vector machine (SVM) classifier to recognize my pinch grips, which I can use for triggering purposes. As one can accelerate anywhere in space, jerk-based triggering can virtually happen at every position. In air performance, where the performer can move in any direction, the relativity of jerk-based excitation may not always be favorable. Hence, for more precision-requiring actions, gesture recognition is crucial when performing based on muscle signals.

Second, I used several chaotic attractors, such as Hénon-and-Heiles or Lorenz systems, to create melodic motives. As mentioned previously, the EMG was pitch-shifted at audio rate using additional oscillators. When using a pinch grip, the SVM model recognizes it and draws a new set of points on the orbit; each point refers to a frequency. Therefore, although the new frequency may sound random compared to the previous one, it converges to a melodic line. However, in practice, that does not always work as expected. For example, if the time interval between two points is too long, it never really converges to a globally familiar pattern. If the interval is too short, on the other hand, it can become too repetitive.

Third, the system has two multi-layer perceptron (MLP) artificial neural networks (ANNs). They can be used both in pre-trained mode or in online training mode. The networks were used with a simple gamification strategy. Each ANN maps eight EMG channels of one armband to a point in an XY plane, of which both axes are mapped into an oscillator parameter. First, in the training phase, you create a sound trajectory as you like, using muscle contractions. That can occur before or during the performance, while the latter makes more sense when playing with an ensemble. Then, these two trajectories (one per forearm) are shown in different colors on the GUI window, with moving circles representing your current mapped motion. As shown in Figure 3.5, the goal of the "game" is to make two points meet so that a new random event is triggered. As a performer, this is one of the fascinating features of the system. However, such a gamified strategy can create a high cognitive load.

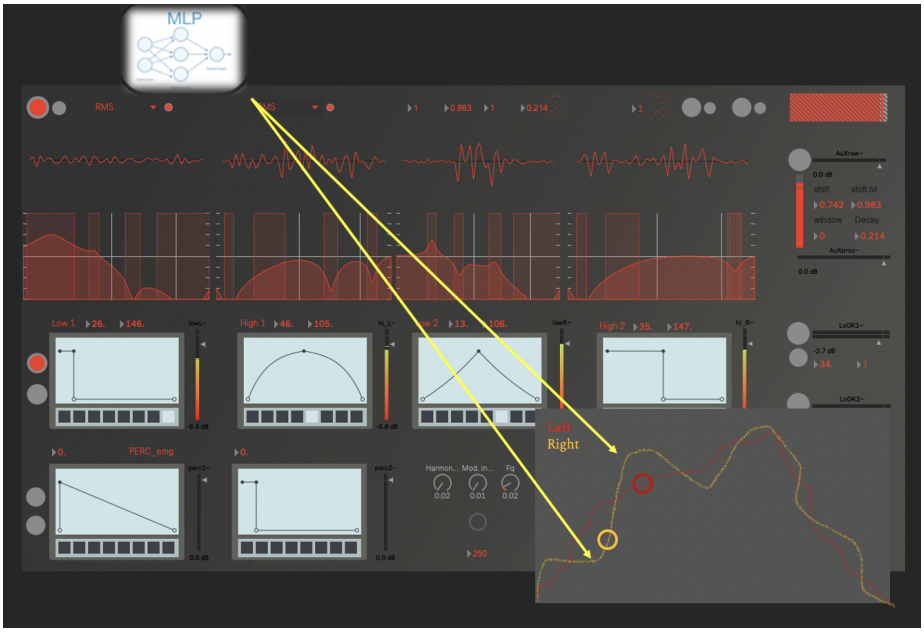


Figure 3.5: Screenshot of the GUI of RAW. The "gamification" window is magnified and displays the nonlinear trajectories for two points. When the red and yellow circles meet, an event is triggered.

RAW is based on real-time audio analyses for automated ensemble interaction. The system relies on the *pipo*~ plug-ins from the IRCAM MuBu library (Schnell et al., 2019). Real-time audio analysis is challenging at many levels, particularly in free improvisation settings. The solution was to use an adaptive algorithm and to limit the scope of the system to rhythm-related tracking using mainly spectral flux and dynamics-tracking using envelope-following. The patch can be adjusted for different ensemble constellations. For example, I used the library's YIN algorithm implementation for monophonic pitch estimation when performing with a vocalist (Figure 3.6).

RAW also incorporates an effects outboard with a selection of time-based processing modules. These can be employed for live sound processing, which can have highly efficient results in duo performance. However, in bigger ensembles, such processing can introduce too much ambiguity.

3.3.3 Updates

As part of the iterative prototyping, newer versions of *RAW* have been developed after the publication of Paper III. The second version of *RAW*as made for a long solo set in a live-streamed festival during the first wave of the pandemic. I usually play sets that last no more than 20-30m, but this time we played for one hour. Here I decided to focus on live sampling making a mashup that could

Figure 3.6: A collage of four performances featuring RAW. From top-left, a duo with a drum set in Istanbul Turkey; a trio performance with live coding and vocal & laptop, in Oslo, Norway; a quintet with live coding, shared electric guitar & laptop, voice & laptop, in Trondheim, Norway; a duo with a gestural controller, in Istanbul, Turkey.

alternate between sound-based and beat-based aesthetics. For that purpose, I prepared different modules that could be controlled simultaneously. These allowed for making layers for different musical elements, such as impulsive, sustained, and textural. In Figure 3.7, the respective modules are *perc1* & *perc2* (a corpus of percussive audio samples), *T-Stretch* (time stretch), and *ZJ*). For the latter, I used *Latent Timbre Synthesis* (LTS) by Tatar et al. (2020), a variational autoencoder (VAE) model to create novel textures based on latent space interpolation of audio samples. LTS comes pre-trained, thus it must be used with its dedicated corpora of a few different musical styles, among which I used samples of electroacoustic composition and contemporary classical music. My plan was to use LTS in real-time and create interpolations by "drawing in the air." However, after initial testing I decided to create an offline sample bank to avoid too much latency in performance.

RAW v1.2 was performed in an online live event and I decided to add a video module that processed the live video streamed from my web camera (Figure 3.8). The audience feedback to this visualization was highly positive. People commented that it provided an additional modality that showed the embodied processes of the performer. This is particularly important in the case of "flat" live stream concerts. However, watching the video recording afterwards, I observed that such distorted glitch visuals may also block the visual connection

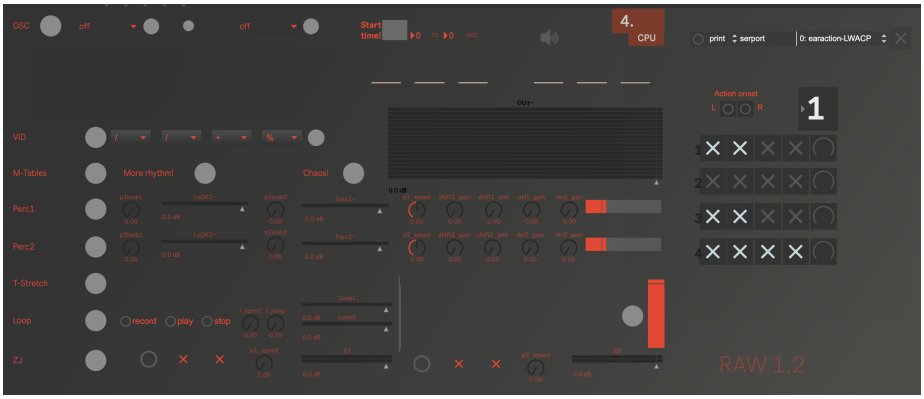


Figure 3.7: The GUI of *RAW* v1.2, featuring a set of new live sampling, looping, live visuals, and foot switch presets.

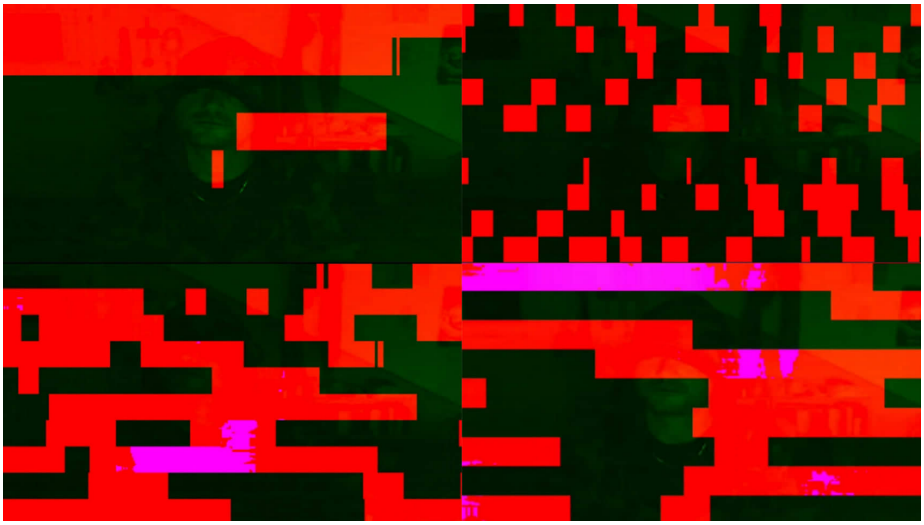


Figure 3.8: Screenshots from a live video of performing with *RAW* 1.2. The video effects processing is mapped to the muscle energy envelope (RMS) to reflect the performer's muscle contraction and relaxation.

between the performer's actions and the resultant sound. This is something that I would like to explore more in the future.

In the third version of *RAW*, I constrained the design around the "air guitar" concept. The aim was not to re-create the guitar, but to re-use the embodied knowledge of it, which Magnusson (2019) describes as *ergomimesis*. To that aim, I first reduced the wavetable buffers from four to two and used the right forearm muscles. That way, I could emphasize the common excitation action aspect of the right forearm. Then I added pitch banks, the scales of which can be selected

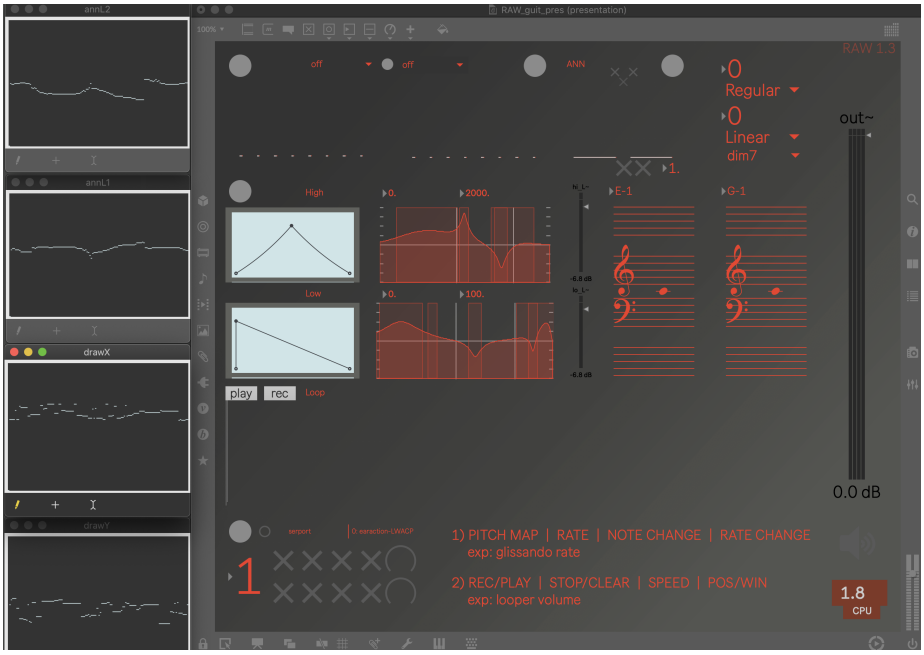


Figure 3.9: The GUI of*RAW* v1.3. The significant change in this version was a move towards an "air guitar" concept and apply the ANNs from the previous "gamification" strategy into a contour-based note selection system. The top two tables on the left represent the trajectories of the learned motion, and the bottom left ones are reserved for pre-defined note trajectories. The scales can be selected, randomized, or automated on the right panel.

by the performer, randomized, or automated. The ANNs I used for gamification in the published version, this time, were used to map the left forearm motion and muscle activation to trajectories for note selection. That is, the performer can change pitches using wrist-extension to go up and wrist-flexion to go down within trajectories (Figure 3.9. This version also relied on using a footswitch for changing the presets, such as the note scale.

3.4 Playing in the Air

The modifications that went into *RAW* v1.3 inspired a new project on guitar ergomimesis. Magnusson (2019, p. 36) suggests this term for mimicking the *ergon*, Greek for work or function. Thus, ergomimesis denotes carrying out the function and the incorporated working memory, *ergogenetic* memory, from one context or domain into another. I began from an "air guitar" perspective, although the aim was never to mimic the guitar in the air. Instead, I wanted to employ the embodied knowledge of playing a guitar and used these possibilities and constraints in the construction of a new instrument.

The first part of the project involved a controlled experiment in a laboratory context. This was entirely different from my previous projects, which had began from my own artistic expliration. The methodological framework can be described as follows:

- Collecting a multimodal dataset of EMG and motion capture data, and video and sound recordings.
- Testing some conceptions regarding the functional categories of music-related motion (see Section 2.3.4), with a particular focus on how biomechanical muscle signals transform into sound in playing the guitar.
- Exploring modeling approaches using the collected dataset for designing new musical interactions.

3.4.1 Data Collection

We recruited participants through an online invitation published on a specified website of the University of Oslo, Norway, and also announced the experiment in various communication channels. Participation was rewarded with a gift card (valued approximately 30). Such a recruitment method had some consequences. The diversity of participants was limited to whoever volunteered. Unfortunately, we only had one female participant. One can reasonably argue that as an obstacle for generalizing the statistical results. Another limitation was the experimental setup in a controlled laboratory environment, which felt unnatural to several of the participants. In addition, the recruitment award did not appeal to professional musicians. The thirty-six participants who took part in the study were primarily semi-professional musicians and music students. Before conducting the research, we obtained ethical approval from the Norwegian Center for Research Data (NSD), Project Number 872789. All recording sessions took place in the *fourMs* lab at RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo.

Motion Capture

While *motion capture* covers a variety of techniques to record motion data, it is usually referred to as passive marker-based infrared (IR) motion capture (MoCap) systems (Jensenius, 2018a). These systems can provide high spatiotemporal resolution, thus their use is common in music-related motion research (Perez-Carrillo et al., 2016; Kelkar, 2019), including micromotion (Gonzalez-Sanchez et al., 2018), interaction (Skogstad, 2014), and modeling (Caramiaux et al., 2012; Wallace et al., 2020). A combination of optical body-tracking with physiological sensors that capture "covert" information provide rich data of both kinetics and kinematics.

We recorded overt upper-body motion in this study with 12 optical cameras capturing at a frame rate of 200 Hz, using a Qualisys Oqus system. We preferred

the Qualisys system above an Optitrack system also available in the lab because of Qualisys' AIM (Automatic Identification of Markers) model which facilitated marker labeling. It was also straightforward to synchronize the MoCap recording with our Delsys EMG system. It was not possible to easily record Myo data in sync with the motion capture. So we had to develop a custom-built software solution (**?**) for synchronizing the signals. In the end, we had a fully mobile system that can also be used for recording synchronized motion–sound data virtually anyywhere.

Our aim in this study was to re-use the learned action repertoire of guitar performance in a new context. In *RAW* v1.3, I simulated the excitation action by using the acceleration signal from the IMU unit worn on the right forearm. A fluently working model can be easily implemented with an elaborate combination of acceleration and orientation, so no learning algorithm is necessary for that purpose.

Muscle-Sensing

The excitation action on the guitar that can be captured as overt motion in space reflects what would be seen in the mirror, *le corps objectif*, using a term by Merleau-Ponty (2012). According to that, then, one would look at what is somewhat covert inside the "living body" (*le corps propre*) if the aim was to find something unique in a typical action repertoire. Most people would have an idea about how a sound-producing excitation action looks like on the guitar. Is that also the same for internal bodily processes? In other words, muscle activations, do they follow the same trend of actions as seen from outside?

Electromyography (EMG) is the technique of measuring the electrical activity produced by muscles (Phinyomark et al., 2020). In the user study of a muscle-based sound effects controller I developed (Erdem et al., 2017) before my dissertation project, I observed that muscle signals while playing the guitar were not following the trend of overt actions in a linear fashion. Many participants in that study reported enjoyment of the unexpected responses of the device, which positively dragged them into a more exploratory musical approach. That system was based on the mechanomyogram (MMG), mechanical signals effectuated by contractions in muscle fibers (Watakabe et al., 2001). In the operative order of the human body, EMG measures the electrical nerve stimulation that contracts muscles resulting in the MMG. Tanaka (2015, p. 1) describes it as: "The EMG is not an external sensor reporting on the results of a gesture, but rather a sensor that reports on the performer's *intention* to make a gesture." Wearing two light-weight and stable consumer-grade products, such as the Myo armband, it is possible to obtain sixteen bioelectric channels representing the muscle groups surrounding both forearms responsible for the excitation and modification actions.

Pérez (2010, p. 3) asks if there is "something inherently 'musical' in the patterns that we can observe" in biological signals. Performing with RAW, I found that the extreme signal peculiarities of EMG and the unconventional control they enabled were indeed musical at many levels. In this study, however,

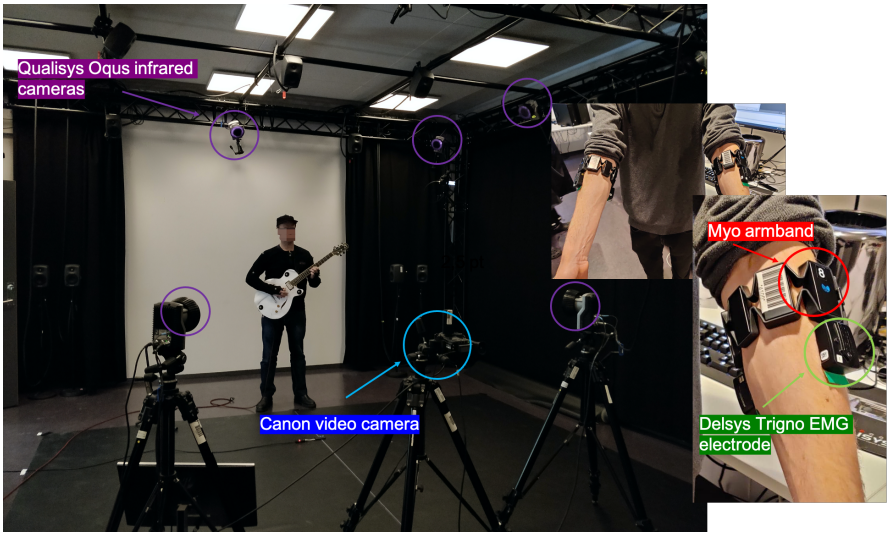


Figure 3.10: Images from one of the experiment sessions. Notice some of the infrared motion capture cameras hanging from the ceiling and some standing on tripods, a video camera, and the placements of two different EMG system electrodes. The monitor with instructions for the performer can be seen below the front left motion capture camera.

I wondered to what extend the mimesis of guitar-playing can bring that noisiness to an equilibrium. Making new instruments and performing with them yield fascinating subjective experiences. Now I was curious about how these experiences change among musicians.

In the experiment, we decided to combine the Myo armbands with a medical-grade *Delsys Trigno* EMG sensor system. The latter provides high-quality data suitable for analytical purposes (see Pizzolato et al. (2017), for a comparison of various EMG acquisition setups). While the Myo sensors can acquire EMG at a frame rate of 200 Hz, we recorded the raw EMG data at 2000 Hz using the Delsys system. The logic behind such a combination was to use the Delsys data for analysis and reserve the Myo data for developing the interaction model. As for the sensor placement, we followed the exact forearm location used in both Vrengt and RAW: We placed two Delsys EMG sensors on each side of the forearm corresponding to the extensor carpi radialis longus and flexor carpi radialis muscles, just below the Myo armbands. That protocol was near-optimal for interactive music control and for the training dataset we wanted to collect. During the analyses, however, we realized that additional electrodes placed on the upper arm would be helpful in particular to compare the muscle activations during tasks that required challenging agility. Figure 3.10 shows the placement of the EMG electrodes together with the infrared and video cameras.

3.4.2 From EMG to Sound on the Electric Guitar

As elaborated in Section 2.3.5, temporal and dynamic similarity is crucial for the perception of causality in music performance. Therefore, the analyses focused on similarities between the EMG RMS of each of the four channels (two per arm) and the sound RMS for each participant. The functional motion types (impulsive, iterative, sustained) and the time scales (sub-chunk, chunk, supra-chunk) I presented in Section 2.3.4 provided the theoretical apparatus when designing the experimental tasks. While the given tasks were based on the guitar-like versions of these fundamental motion types, we also recorded free improvisations for modeling and exploring coarticulated patterns.

The statistical analysis focused on Pearson's product-moment correlation, Spearman's rank correlation, and analysis of variance. The results showed a significant correlation between muscle activations and the resultant sound energy envelope in playing impulsive tasks. Even though one can observe common patterns among different players' data in iterative tasks, the muscle–sound similarity was not statistically significant. Emerging patterns during modification actions of the left forearm muscles were the most interesting, in my opinion. Even in the simple task of sustaining a single note for a few seconds, it is fair to say that each player demonstrated somewhat unique muscular patterns. Recalling the "ecological approach" from Section 2.3.2, these peculiarities can be seen as potential affordances of a new ergomimetic instrument.

The results were satisfactory overall. However, we observed many nonlinearities between the EMG and audio signals. Thus, the chosen statistical methods for correlating bodily signals with sound features remain an open question. To demonstrate such non-linearities, we applied time-varying Principal Component Analysis (PCA) (Santello et al., 2002) to merge the 4-channel data of both forearms to explore prominent features. The input matrix for the PCA was defined as $A \in \mathbb{R}^{m \times n}$ where m is the number of participants and n denotes the number of EMG channels. We obtained two principal components, separately for actions with soft and strong dynamics, as shown by the following equation:

$$EMG_{\mathrm{m}} = \mathrm{meanEMG}_{\mathrm{m}} + PC1 \times EMG1_{\mathrm{m}} + \ldots + PCn \times EMGn_{\mathrm{m}} \quad (3.3)$$

We then combined PCA with Singular Spectrum Analysis (SSA) for further signal–noise separation. Using these tools, we could observe and demonstrate the varying level of non-linearities of muscle–sound relationships for the tasks played at different dynamic levels. Here I should note that, differently from IR MoCap data in music-related motion research, there are no generalized approaches to investigating muscles in music performance. Hence, many aspects of the study were highly exploratory. For example, there is no universal method to find an optimal SSA window length L. Thus, we relied on a rule suggested by Khan & Poskitt (2013), as $L = (\log N)^c$ with $c \in (1.5, 3.0)$ for assigning a window length. Starting from there, as the RMS segments of our interest were at a fixed length of $N = 344$, we empirically chose $c = 2.5$, which yielded to $L = 10$.

In the end, we did not include the camera-based motion data in the published paper. This would have been interesting from an analytic perspective. However,

my intention was to use the data in the development of a new air instrument. The Myo is a much more portable device, hence the EMG and IMU data were better suited for the development part. So we primarily used the motion capture data and video recordings as to clarify ambiguities in the EMG data. For example, the sparse optical flow that we extracted from the video recordings using the Musical Gestures Toolbox (Jensenius, 2018b) helped us to observe the participant's ancillary motion, which was not clear for the naked eye. We could then understand better some unexpected iterative patterns in the data. Beyond that, however, we reserved both video and MoCap recordings for follow-up studies and exploration of new techniques for music interaction.

3.4.3 From EMG to Sound "in the Air"

The functional motion categories provided us with a basis for investigating the motion–sound similarities in playing the electric guitar. Although it is not explicitly mentioned by Godøy (2006), we can think of all possible motion in terms of the coarticulation of impulsive, iterative, and sustained motion chunks. The dataset we collected comprised 248 tasks and 62 free improvisations. The idea was it should be possible to create a predictive model to learn how to map muscle RMS recorded during free improvisations to the RMS of the sound. This we did with EMG data captured by Myo armbands.

For time series prediction tasks, the long short-term memory (LSTM) recurrent neural network (RNN) architecture is a go-to tool (Eck & Schmidhuber, 2002; Martin et al., 2018a). Martin & Torresen (2019) suggests 32 or 64 LSTM units in each layer as the most appropriate for interactive systems. Thus, we trained nine models with one, two, and five hidden layers and each containing 16, 32, and 64 units to test the latency of different configurations. In short, the input to the network was a 16-dimensional array of raw EMG signals, fed into the network as sliding windows of 250ms with the target of a single sample of sound RMS at a time. We defined the training loss function as:

$$\mathcal{L}(x_{\mathrm{RMS}}, \hat{x}_{\mathrm{RMS}}) = \frac{1}{n}\sum_{i=1}^{n}(x_{\mathrm{RMS,i}} - \hat{x}_{\mathrm{RMS,i}})^2\,, \tag{3.4}$$

where x_{RMS} are the recorded values, $\hat{x}_{\mathrm{RMS}}$ are the values to be predicted, and the sliding window has size n.

The results satisfied our expectation such that the model was able to predict the sound RMS of free improvisations, based on a dataset of fundamental motion types.[3] On the other hand, the average latency of even the smallest network configuration was around a quarter of a second, while the best one was reaching a whole. Besides, a muscle-based estimation of sound dynamics could also be made through traditional signal processing techniques. Our modeling approach successfully provided empirical support to embodied music cognition conceptions, as mentioned above. However, for developing an interactive system to perform in real-time, there were more to be done.

[3] A video of the offline sonification test can be found at https://youtu.be/-_wgBZY2iF8

3.5 CAVI

The model we developed in the previous study could associate patterns of muscle and sound data. Unfortunately, it had a considerable latency. As discussed in Section 2.3.5, a number of studies show how participants attribute agency to others than themselves in situations where they experience delay or temporal distortion. One of the takeaways from such experiences is that perceptual cues are as critical as sensorimotor ones. Gurevich et al. (2010) proposes that exploring the constraints of instruments is essential in artistic practice. I therefore decided to build CAVI based on generative modeling and the dataset at hand.

3.5.1 Agent Architecture

In Section 2.2.3, I grouped machine learning (ML) tools based on the intended purposes in music applications. These were mapping, analysis, and generation. While the model we developed in the previous project lays somewhere in between mapping and analysis, CAVI focused on generation. Here the key shift was from a model that learns the discriminative properties of data to a modeling framework that makes predictions by sampling from a probability distribution. While the former learns the boundaries of the data, the latter captures how it is distributed in the data space. Foster & Safari (2019, p. 4) defines such an approach as a probabilistic model that generates an output of "a high chance of belonging to the original dataset." An analogy would be that while one approach predicts the ingredients of a dish, the other tries to re-cook from the taste it remembers. One way of doing that with sequential data is combining a recurrent neural network (RNN) with a mixture density network (MDN) (Bishop, 1994). MDRNNs have over the years proved generative capacity in projects such as speech recognition (Schuster, 1999), handwriting (Graves, 2013), and drawing sketches (Ha & Eck, 2017).

In simple terms, the aim is to add a sampling layer to the output of an LSTM model, such as the one we trained for action–sound modeling to "play in the air." Mixture density networks (MDNs) treat the outputs of the neural network as parameters of a mixture distribution (Ellefsen et al., 2019). That is often done with Gaussian mixture models (GMMs), which are considered particularly effective in sequence generation (Goodfellow et al., 2016, p. 190), and appropriate for modeling musical improvisation processes (Martin & Torresen, 2019). The output parameters are mean, weight, and standard deviation. A GMM can be derived using these parameters of each mixture component (the amount is defined as hyperparameter) and be sampled to generate real-valued predictions.

As depicted in Figure 3.11, CAVI's model consists of an RNN with two layers of LSTM cells (Schmidhuber, 2009). Each LSTM cell contains 64 hidden units, based on the findings from the previous study. The second layer's outputs are connected to a MDN. As our GMM consists of $K = 5$ n-variate Gaussian distributions, each representing a possible future action, the LSTM layers learn to predict the parameters of each of the five Gaussian distributions of MDN. For optimization, we minimize the negative log-likelihood of sampling true values

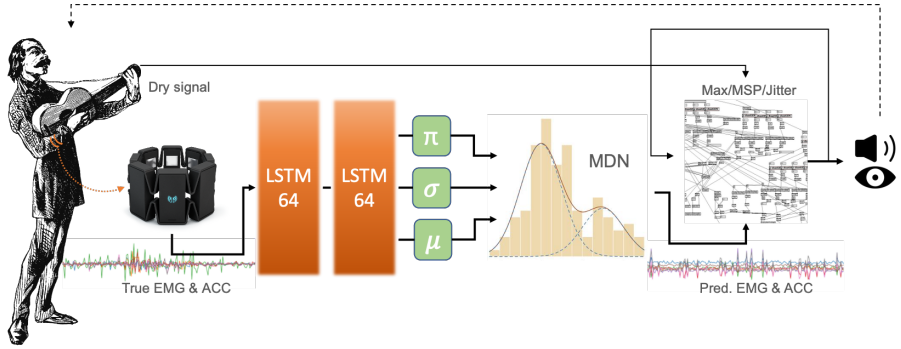


Figure 3.11: A simplified diagram of the signal flow through MDRNN. The model receives EMG & ACC from the Myo armband. The MDRNN outputs the mixture distribution parameters, from which we sample a new window of EMG & ACC data. The generated data is sent to Max/MSP/Jitter, where the visuals are generated, and the dry acoustic instrument sound is processed through several EFX modules. Notice that the Max patch also encapsulates the rule-based structure within which CAVI continuously tracks the audio outputs and makes the necessary adjustments.

from the predicted GMM for each example. A probability density function (PDF) is then used to obtain the likelihood value. For simplicity in the PDF, these distributions are restricted to having a diagonal covariance matrix, and thus the PDF has the form:

$$p(\theta; x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \Sigma_k; x) \tag{3.5}$$

where π are the mixing coefficients, μ, the Gaussian distribution centres, Σ the covariance matrices and n is the number of values corresponding to EMG and acceleration (ACC) data contained in each frame. The Adam optimizer (Kingma & Ba, 2014) was used in the training until the loss on the validation set failed to improve for 20 consecutive epochs. This configuration corresponded to 56331 parameters. The loss is calculated by the *keras-mdn-layer* Python package (Martin & Duhaime, 2019), which makes use of the Tensorflow probability library (Dillon et al., 2017) to construct the PDF. In the generation phase, it was possible to continuously adjust the model's level of "randomness" by tweaking π and σ temperatures. For example, larger π temperature results in sampling from different distributions at every time step.

3.5.2 Composition

Martin (2019) shows how MDRNN can be used in a call-and-response mode. If you train the model with a dataset of your improvised melodies on a keyboard, for instance, it "guesses" how you would carry on with the melody you started

to play and stopped at some point. This is similar to call-and-response systems developed in jazz contexts, such as the *Continuator* of Pachet (2003). The main difference is that Martin (2019) uses a motion dataset; thus, the model generates control signals in response to or as continuation of the user's actions. That echoes, first, how Sawyer & DeZutter (2009) describes improvisation as "collaboratively emergent," and, second, one of the topics presented in Section 2.3.5, which exemplified an expert player who can predict the outcome of another player's actions. Knoblich et al. (2011) defines that as an ability to simulate actions. Such potentially joint situations can make possible what he calls an emergent coordination. In other words, if one predicts the other's actions, and those predictions make sense for the actor, a coordination can emerge.

One interesting question is whether coordination or joint action can emerge between a performer and a musical agent that somewhat simulates the performer's likely actions by means of generative predictions? To explore that, CAVI continuously tracks the performer's motion input, consisting of 4-channel EMG and 3-channel ACC, and generates what will likely come next. I built a custom Python script that runs the model in the background throughout the performance (Erdem, 2021). As one can notice, our dataset from the Myo armbands consisted of data from both forearms. However, in this project, another set of constraints was to exclude the data from the left forearm. The statistical results from the previous study showed significant generalizability only for the data from the right forearm responsible for the excitation actions. As I also touched upon in the last section, the left forearm muscles often exhibit quite peculiar patterns. Thus, following a series of training and test sessions using data from both forearms, we empirically decided to limit CAVI to generate control signals solely based on the performer's excitation actions. This strategy seems to have worked well as both performers in the evaluation stressed the predictability of CAVI's output (Paper V).

Sound

Differently from previous projects presented in this chapter, CAVI followed a musical strategy that focused on live sound processing in duo improvisation. The musicians who tested the system performed on plucked (guitar) and percussive (drums) instruments. During their performance, CAVI continuously generated new EMG and ACC data akin to the musician's excitation actions. The generated data were used as control signals mapped to parameters of digital audio effects (EFX) modules. This could be seen as playing the electric guitar through some EFX pedals while someone else is tweaking the knobs of the devices.

CAVI's EFX modules primarily rely on time-based sound manipulation, such as delay, time-stretch, stutter, etc. The jerk of the generated ACC data triggers the sequencer steps (Figure 3.12, which functions as a matrix that routes the EFX sends and returns. Depending on user-defined or randomized routing presets, the EFX modules are activated by the trigger the model generates. The generated EMG data (corresponding to the same flexion and extension muscle groups similar to previous projects) is mapped to EFX parameters. The

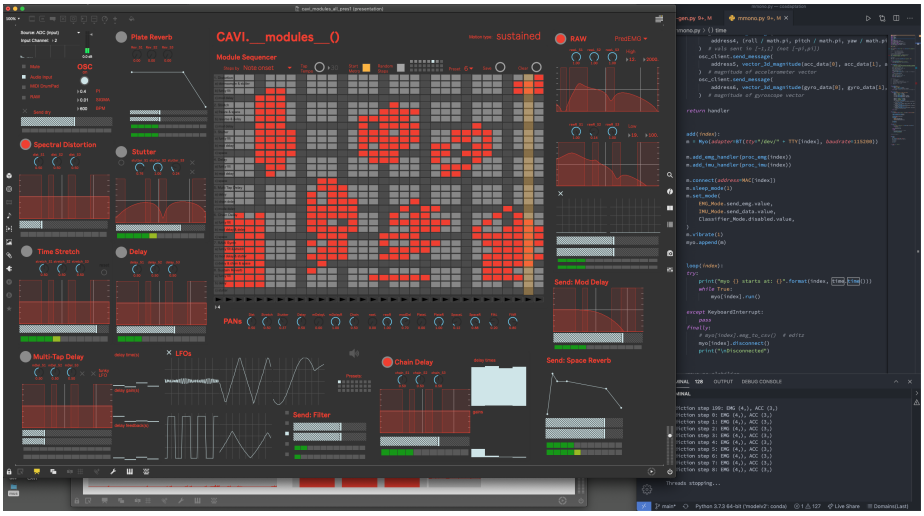


Figure 3.12: A screenshot of CAVI's "cockpit" for inspecting the automated processes throughout the performance. The Max window is on the left and contains a grid where shapes can be drawn to determine the overall compositional structure. Notice the EFX modules surrounding the grid, each having individual send/return for interconnection with other modules. The Python script on the right is continuously retrieving data from the Myo armband, pre-processing and windowing, feeding it into the model, and finally streaming the generated data through OSC to the Max patch.

real-time analysis modules track the musician's dry audio input and adjust EFX parameters according to pre-defined thresholds. These machine listening agents include trackers of onsets and spectral flux. For example, if the performer plays impulsive notes, CAVI increases the reverb time drastically, such that it becomes a drone-like continuous sound. If the performer plays loudly, CAVI decides about its dynamics based on the particular action type of the performer (see Section 2.3 for more details about such action types).

The strategy implemented in CAVI has some significant drawbacks. Firstly, the model architecture was not suitable for multimodal data. One can observe how generated ACC and EMG data influenced each other. Second, live EFX is not aesthetically favorable for everyone, something one of the musicians who performed with CAVI clearly stated in his reflection. Third, ambiguity was often too high compared to a strategy where CAVI has an entirely different sound palette than the musician.

Visual

CAVI is an audiovisual instrument not only for aesthetic reasons but also to relieve potential causality ambiguities. The "body" of the virtual agent is a

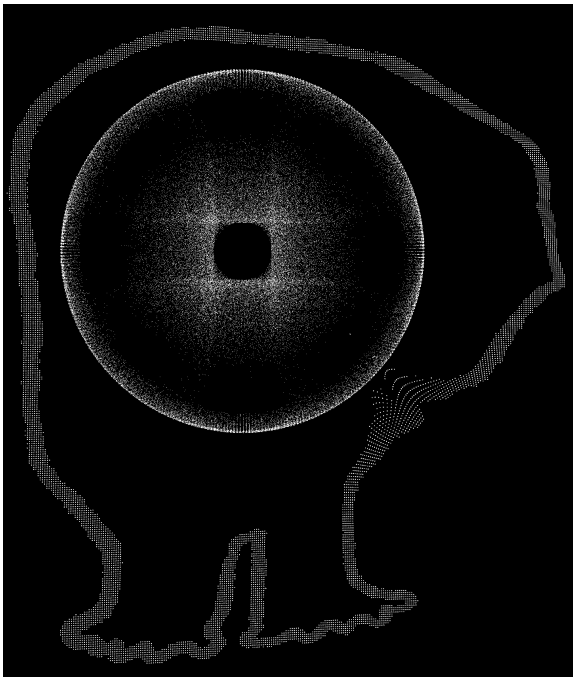


Figure 3.13: CAVI's virtual embodiment. The body contours are hand-drawn by Katja Henriksen Schia. The eye is developed in *Max/Jitter* using *OpenGL*. It is based on two layers of *jit.matrix*: The first matrix contains the digitized pixels of the virtual body shape. The second matrix encapsulates $350 * 350$ particles on a two-dimensional plane

digitized version of a hand drawing by Katja Henriksen Schia. CAVI's "eye" is designed in Max/Jitter using OpenGL as shown in Figure 3.13. The design aims at presenting CAVI as an uncompleted, creepy but cute creature that has only legs that are too small for its body, no arms, a tiny mouth, and a big eye. In the real-time animation, the body contracts but does not make full-body gestures. Instead, the eye blinks from time to time when CAVI triggers a new event, opens wide when the density of low frequencies increases, or stays calm according to the overall energy levels of sound.

One of the studies presented in Paper VI, investigated whether ascribed agency promote engagement in interactive art, using a similar eye concept as the basic visual component. In the study, visitors interacted with a browser-based widget that tracks the mouse cursor in four conditions. The findings suggested that visitors attributed the most agency to the *angular offset* condition in which the eyes of the agent, *Dot*, were offset in the direction of the cursor plus some angle that drifts over time using Brownian motion. This suggests that people favored the aspect of surprise in the interaction. CAVI is built with similar unexpected, yet controlled moves.

3.5.3 Performance

The CAVI paper included in the dissertation (Paper V) primarily focuses on an evaluation of the system with two expert performers. In the previous projects, I always performed on the systems I developed. In this project, on the other hand, I did not actively perform and rather observed and interviewed performers (a guitarist and a drummer) and collected audience responses to a questionnaire. We also collected quantitative data during the performance, such as EMG, IMU, and breathing, yet reserved it for future study. We took many critical decisions regarding the performance organization, a considerable amount of which were either coincidental or bound upon some limitations. From the selection of invited musicians to organizational details, such as what was written in the flyer, can influence the outcome. For example, recalling the discussion of Strasser (2015) in Section 2.2.6 about how specific requirements from artificial agents can be way too demanding, the expectations of both the performers and some audience members from a system that was promoted as AI seemed to be also high. That is also in line with some criticism regarding the overestimation of AI's creative agency (Dahlstedt, 2021). In addition, we did not communicate with the musicians enough regarding the system details. For example, they explained in their post-concert interviews that they had expected a fully functioning free improvisation agent. However, what I had in mind was a designed improvisation, or a "composed improvisation" (Zicarelli, 1987).

CAVI is still developing and learning, and its capabilities can at the moment best be described as a "musical AI toddler." Its emerging human-machine interactions cruise on the limits between enriching vs. competing. The main drive is to challenge the guitarist's embodied knowledge and musical intentions. The performance showcased CAVI's artistic stance, arguing that interaction can also be an aesthetic choice as part of a composition, as much as one makes decisions about, for example, the sound generation or harmony. As it turned out, we did not focus enough effort when preparing the physical space (which can be seen in Figure 3.14) for the performance. The acoustics of the space, combined with some sound system issues, negatively influenced the performer's on-stage experience. In addition, it also increased the temporal ambiguity that CAVI was already introducing. Both musicians reported that their experience in listening back to the recording of the performance was radically better than their live experience.

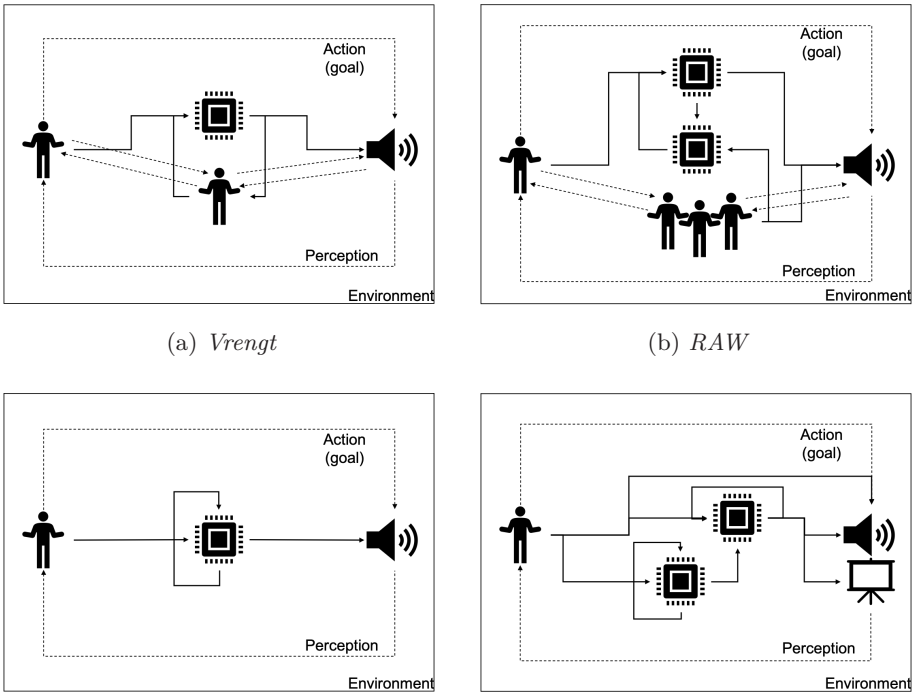
3.6 Summary

In this chapter, I presented my iterative methodology. I have employed a number of different methods throughout my PhD research, many of which were concerned with the use of technology. I grouped these methods under two overarching titles, *interaction* and *evaluation*, in Table 3.1. These were further sub-grouped in terms of prominent parts of interactive systems (sensors, control/mapping, prototyping, and performance) and evaluation studies (data collection and analysis). The first two projects (*Vrengt* and *RAW*) relied on exploring different control approaches,

Figure 3.14: The stage where CAVI's premier took place at the Science Library, University of Oslo. (Photo: Alena Clim)

which were evaluated through self-reports and subjective analyses. The third project ("air instrument") provided statistical results and collected a multimodal motion–sound dataset. The last project (*CAVI*) was built on an accumulated knowledge and data and focused on an ecological evaluation. The performance diagrams of the four projects are depicted in Figure 3.15, which demonstrate how performers and machine(s) are situated within the environment.

Among the four projects, I took part in two of them as a performer. Hence, the evaluation in those studies relied on self-reports and subjective analysis. The third project was a statistical study and did not include an evaluation as such. In the last project, I had an observer role, which was different than the previous ones. These alternative perspectives helped me realize an important dimension of working with music technological tools. That is, developing an interactive music system is inherently different from building acoustic instruments. The programmer/developer inevitably becomes the composer of the system, echoing the notion of "composed instruments" (Schnell & Battier, 2002). For example, when I performed with CAVI, I realized that I built it for myself. It is crucial to communicate the intentions of a system with other performers. Otherwise, the composed aspect of a system can negatively impose the programmer's musical choices on other musicians. All in all, I have learned a lot by combining subjective, statistical, and observational perspectives. I still have a long way to go, but the iterative process has helped in moving the various projects further.



(a) *Vrengt* (b) *RAW*

(c) *Playing "in the Air"* (d) *CAVI*

Figure 3.15: A collage of simplified diagrams illustrating the variation of approaches to constructing feedback pathways in four interactive systems developed as part of this dissertation.

Chapter 4

Research Summary

Our greatest glory is not in never falling,
but in rising every time we fall.
– Confucius

4.1 Introduction

In this chapter, I will present and discuss the six papers that are included in the dissertation. They are included in chronological order, but as a natural consequence of the practice-based and iterative research process presented in the previous chapter, the research project did not proceed linearly. Instead, concepts and methods emerged throughout the realization of the four projects that constitute the basis for the six papers. For example, the evaluations made as parts of Papers I and III led to Paper IV's research design, for which Paper II provided a conceptual and terminological basis. Similarly, Paper V reports the evaluation of a project built upon previous findings and methods, and, in tandem, part of Paper VI investigated a core concept that emerged from that project.

4.2 Papers

4.2.1 Paper I

Reference: Erdem, Ç., Schia, K. H., & Jensenius, A. R. (2019). Vrengt: A Shared Body–Machine Instrument for Music–Dance Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 186–191). UFRGS.

Abstract

This paper describes the process of developing a shared instrument for music–dance performance, with a particular focus on exploring the boundaries between standstill vs motion, and silence vs sound. The piece *Vrengt* grew from the idea of enabling a true partnership between a musician and a dancer, developing an instrument that would allow for active co-performance. Using a participatory design approach, we worked with sonification as a tool for systematically exploring the dancer's bodily expressions. The exploration used a "spatiotemporal matrix," with a particular focus on sonic microinteraction. In the final performance, two Myo armbands were used for capturing muscle activity of the arm and leg of the dancer, together with a wireless headset microphone capturing the sound of breathing. In the paper we reflect on multi-user instrument paradigms, discuss our approach to creating a shared instrument using sonification as a tool for the sound design, and reflect on the performers' subjective evaluation of the instrument.

Discussion

In this project, I aimed at developing an interactive system for the co-performance of a dancer and a musician. Unlike many interactive dance systems, we wanted the musician and the dancer to control the same musical parameters instead of working in separate layers. The motivation behind that was to explore the musical possibilities gained by exploiting the complex relationships between multiple agents. In this context, the concept of *waiving the control* was essential. We focused on muscle-sensing to capture the human micromotion. The work that resulted through a participatory design approach was a three-part comprovisation piece that has so far been performed on stage three times.

Following the performances, [1] the evaluation focused on performers' self-reports. The musician reported that the other agent's body and the machine's data processing and sound generation abilities enacted his presence. That echoed the notion of "shared control" discussed in AI and robotics and portrayed a sense of agency distributed among the dancer, the machine, and the musician. The dancer reported an alteration in her sense of agency and body-awareness that she described as a "new type of body." Another critically emphasized aspect was the notion of uncertainty and surprise. "A state of not knowing where to, and how to," the dancer described how she approached performing with the system. The project's primary outcome was evaluating the experience of a form of co-dependency and embodiment among multiple agents. The use of physiological signals as the main interaction channel afforded an embodied performance between intentional versus unintentional. Sharing the control was a significantly different experience than conventional action–sound mappings. The project focused on humans' shared performance and opened for further exploration in Papers III & IV & V.

[1] Video available at https://youtu.be/hpECGAkaBp0

4.2.2 Paper II

Reference: Jensenius, A. R., & Erdem, Ç. (2022). Gestures in Ensemble Performance. In R. Timmers, F. Bailes, & H. Daffern (Eds.), *Together in Music: Coordination, expression, participation* (pp. 109–118). Oxford University Press.

Abstract

The topic of gesture has received growing attention among music researchers over recent decades. Some of this research has been summarized in anthologies on "musical gestures," such as those by Gritten and King (2006), Godoy and Leman (2010), and Gritten and King (2011). There have also been a couple of articles reviewing how the term gesture has been used in various music- related disciplines (and beyond), including those by Cadoz and Wanderley (2000) and Jensenius et al. (2010). Much empirical work has been performed since these reviews were written, aided by better motion capture technologies, new machine learning techniques, and a heightened awareness of the topic. Still there are a number of open questions as to the role of gestures in music performance in general, and in ensemble performance in particular. This chapter aims to clarify some of the basic terminology of music- related body motion, and draw up some perspectives of how one can think about gestures in ensemble performance. This is, obviously, only one way of looking at the very multifaceted concept of gesture, but it may lead to further interest in this exciting and complex research domain.

Discussion

In this book chapter, we wanted to clarify the terminology and provide an overview of fundamental concepts regarding music-related body motion and discuss them primarily in the context of ensemble performance. First, we elaborated on how motion, a physical term for displacement, becomes an action, a rather psychological constitute when directed by a goal. Then, following a brief presentation of functional categories, we suggested that actions that do not have a communicative meaning, such as a pianist hitting a key with the finger, are not necessarily gestures. Hence, meaning is essential to the term gesture. In the second part of the chapter, we discussed four fundamental dimensions of music ensembles and how they influence the way musicians gesture: (1) Ensemble size and setup, (2) musical degrees of freedom, (3) musical leadership, and (4) machine musicianship. The chapter constitutes an important aspect of the theoretical development of this dissertation work.

4.2.3 Paper III

Reference: Erdem, Ç., & Jensenius, A. R. (2020). RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 477–482). Birmingham City University.

Abstract

This paper describes the ongoing process of developing *RAW*, a collaborative body–machine instrument that relies on 'sculpting' the sonification of raw EMG signals. The instrument is built around two Myo armbands located on the forearms of the performer. These are used to investigate muscle contraction, which is again used as the basis for the sonic interaction design. Using a practice-based approach, the aim is to explore the musical aesthetics of naturally occurring bioelectric signals. We are particularly interested in exploring the differences between processing at audio rate versus control rate, and how the level of detail in the signal—and the complexity of the mappings—influence the experience of control in the instrument. This is exemplified through reflections on four concerts in which *RAW* has been used in different types of collective improvisation.

Discussion

This paper resulted from the second main project of this dissertation. In Paper I, the dancer stressed the intimacy and dynamic range of muscle-based interaction and indicated that the surprise component is essential to the collaborative interaction concept presented in the paper. Her emphasis echoed conceptions of collective improvisation. For example, Sawyer & DeZutter (2009) suggest that the creative agency of an improvisation ensemble is collaboratively emergent. That emergent nature is exploited through unpredictable and coherent processes. From such a perspective that links the concepts of embodiment, surprise, and agency, I sought algorithmic approaches that focus on creating unconventional control structures that exhibit a so-called chaotic behavior. To that aim, in addition to using raw bioelectric muscle (EMG) signals at the audio rate as part of the sound synthesis, the implemented methods included real-time audio analysis for the ensemble interaction, nonlinear differential equations, classification algorithms, and artificial neural networks (ANNs) for generating control signals.

RAW has been performed in four public performances, [2] each with a different ensemble. The mean amplitude of muscle signals was highly unpredictable, possibly due to changes in psychophysiological conditions from one day to another. That was both favorable and not, depending on the musical situation. Even though the surprising aspects were promising and engaging, some amount of predictability was necessary. A meaningful relationship between action and sound, repeatable playing techniques, and structuring the time are critical. For example, unpredictable processes were more favorable when shorter in duration. Group synchrony in larger musical idea spaces required more control. Similarly, performing with a drummer necessitated better control, particularly in time-sensitive triggering actions. All in all, the findings revealed a need for a better understanding of action–sound couplings in traditional acoustic instruments and seeking a delicate balance between control and noise.

[2] Videos available at http://bit.ly/raw_videos

4.2.4 Paper IV

Reference: Erdem, Ç., Lan, Q., & Jensenius, A. R. (2020). Exploring relationships between effort, motion, and sound in new musical instruments. *Human Technology* (pp. 310–347).

Abstract

We investigated how the action–sound relationships found in electric guitar performance can be used in the design of new instruments. Thirty-one trained guitarists performed a set of basic sound-producing actions (impulsive, sustained, and iterative) and free improvisations on an electric guitar. We performed a statistical analysis of the muscle activation data (EMG) and audio recordings from the experiment. Then we trained a long short-term memory network with nine different configurations to map EMG signal to sound. We found that the preliminary models were able to predict audio energy features of free improvisations on the guitar, based on the dataset of raw EMG from the basic sound-producing actions. The results provide evidence of similarities between body motion and sound in music performance, compatible with embodied music cognition theories. They also show the potential of using machine learning on recorded performance data in the design of new musical instruments.

Discussion

This paper is based on a controlled laboratory experiment. We collected a multimodal dataset of EMG, motion capture data, and video and sound recordings of guitarists playing a set of given tasks and free improvisations. The main objective was to understand better the relationships between the temporal shape of an action and its resultant sound and whether these relationships can be used to create action–sound mappings in new instruments. Most studies on music-related motion have focused on overt motion features. As such, it has also been common to create action–sound mappings based on those features. However, the relationships between covert muscle signals and the resultant sound are relatively underexplored. Conceptually, the premise was that these relationships are measurable aspects of acquired skills of playing a traditional musical instrument, which Smalley (1997) explained as an intuitive knowledge of action—sound causalities in traditional sound-making. Starting from there, we first performed a statistical analysis and then used deep learning to develop an action–sound model.

A total of 36 participants performed tasks based on guitar-like versions of each of the three basic sound-producing action types: impulsive, iterative and sustained. The final dataset consisted of 31 participants following the exclusion of the incomplete data. We developed custom Python scripts for capturing, synchronizing, preprocessing and analyzing the data. The statistical techniques employed included correlation coefficients, analysis of variance, Principal Component Analysis (PCA), and Singular Spectrum Analysis (SSA).

The results showed explicit action–sound correspondences, compatible with theories of embodied music cognition. These correspondences' statistical levels depended on the given task and dynamics. First, we found a significant variance when comparing attacks with soft and strong dynamics. When we looked at the sound's frequency spectrum, stronger dynamics led to a brighter sound. Second, more manageable tasks, such as impulsive, yielded a higher temporal correlation. In contrast, we observed how varying levels of motor control of each participant resulted in peculiar EMG waveforms when playing the iterative tasks. Here, particular ways of using rhythms and structuring musical time had a determinant role in the muscle activations. Thus, we argued that complex rhythms yield unique bodily patterns.

Following the empirical exploration of how biomechanical energy transforms into sound, we used these transformations as part of a machine learning (ML) framework based on Long Short-Term Memory (LSTM) networks and compared nine model configurations. The aim was to find out how much latency these models would be subject to when used as part of a musical instrument. In the training dataset, we solely used the performance data of the given tasks. Our results showed that the models could predict audio energy features of free improvisations on the guitar, relying on an EMG dataset of three distinct motion types.[3] However, even the smallest model configuration demonstrated a perceptible latency compared to acceptable ranges (20–30 ms) for real-time audio applications (Lago & Kon, 2004). Such a caveat could be considered as a problem to be solved or be approached creatively.

4.2.5 Paper V

Reference: Erdem, Ç., Wallace, B., Glette, K., & Jensenius, A. R. (2021). Tool or Actor? An Evaluation of a Musical AI "Toddler" with Two Expert Improvisers [Manuscript submitted for publication]. *Computer Music Journal*.

Abstract

In this paper we introduce the coadaptive audiovisual instrument CAVI. This instrument uses deep learning to generate its control signals based on muscle and motion data of the performer's actions. The generated control signals automate the live sound processing based on layered time-based effects modules. How is such an instrument perceived by the performer? Is it an instrument or an actor? We report on an evaluation of CAVI and its use in a public event with two expert improvisers. The evaluation is based on interviews with the performers and questionnaires filled out by audience members. The analyses showed that whether such an instrument is experienced as a tool or actor is closely linked with the performer's sense of agency, which varies throughout a performance depending on several factors, such as perceived qualities of the musical coordination, a delicate balance between surprising and familiar elements, and physical aspects of the performance environment.

[3] Video is available at http://bit.ly/air_guitar_smc

Discussion

This project originated in the concepts of emergent coordination (Knoblich et al., 2011), collaborative emergence (Sawyer & DeZutter, 2009), and temporal (un)predictability (Haggard et al., 2002). Following the considerable amounts of latency that the trained models were subject to, I focused on generative modeling in the second iteration of the same approach. Instead of a discriminative supervised model, I used a recurrent neural network (RNN) combined with a mixture density network (MDN) layer (Bishop, 1994), forming an MDRNN. This model continuously tracks the data streamed from a Myo armband worn on the right forearm of the performer and generates new EMG and ACC data. The generated control signals are mapped to modules implemented in Max/MSP/Jitter. The audiovisual program live-processes the performer's acoustic instrument sound in the performance setup and generates an animated virtual body that represents the artificial agent. During the premiere, I collected audience responses to a questionnaire. Then, I interviewed both musicians (a guitarist and a drummer) who performed with the system. The analysis revealed the importance of surprise, the challenges with the environment (physical space), and the multimodality of musical interactions.

4.2.6 Paper VI

Reference: Krzyzaniak, M., Erdem, Ç., & Glette, K. (2022). What Makes Interactive Art Engaging? *Frontiers in Computer Science*, 4.

Abstract

Interactive art requires people to engage with it, and some works of interactive art are more intrinsically engaging than others. This paper asks what properties of a work of interactive art promote engagement. More specifically, it examines four properties: 1) the number of degrees of freedom in the interaction, 2) the use of fantasy in the work, 3) the timescale on which the work responds, and 4) the amount agency ascribed to the work. Each of these is hypothesized to promote engagement, and each hypothesis is tested with a controlled user study in an ecologically valid setting on the internet. In these studies, we found that more degrees of freedom increases engagement; the use of fantasy increases engagement for some users and not others; the timescale surprisingly has no significant on engagement but may relate to the style of interaction; and more ascribed agency is correlated with greater engagement although the direction of causation is not known. This is not intended to be an exhaustive list of all properties that may promote engagement, but rather a starting point for more studies of this kind.

Discussion

I contributed to the development and design of one of the four studies presented in this paper. In that study, the first author designed a browser-based widget that contained a virtual agent with varying action capacities depending on the

particular test conditions. The main objective of the study was to test whether ascribed agencies promote engagement in interactive art. We defined four test conditions: (1) *Control* condition where two eyes are seen static; (2) the *two eye* condition where two eyes follow the mouse cursor of the visitors; (3) the *one eye* condition, which had the same motion properties as the second condition; and (4) the *angular offset* condition where the position of two eyes are offset in the direction of the cursor plus some angle, which drifts over time using Brownian motion. The findings suggested that visitors ascribed the most agency to the latter (*angular offset*) condition. Though my overall contribution to this paper is relatively small, this project provided me with an opportunity to test the relationship between perceived agency and surprise. In doing so, we used a widget with eyes as the main visual component, similar to the virtual embodiment of the musical agent I developed for Paper V.

4.3 Related Artworks

Throughout my PhD fellowship period, I took part as a developer and performer in several artistic projects closely related to my research. Since these have all been vital in shaping both my theoretical and practical perspectives, I will mention them briefly in the following sections.

4.3.1 Installations

Self-playing Guitars

I developed the audio program for autonomous augmented guitars that interact with each other and users. The public performances and exhibitions included:

- An installation at Life science light event in the Botanical Garden (2019)[4]
- A performance in Tampere, Finland (2019)[5]
- An interactive online installation, *Strings On-Line* (2020).[6] The self-playing guitars and rhythm-playing robots can be seen in Figure 4.1.

Interactive Rhythmic Robots

At the International Conference on Live Interfaces in Trondheim, Norway (2020), we made an interactive installation of a swarm of rhythmic robots (Krzyżaniak, 2021) controlled by “air guitar” gestures.

[4] A video of the Life science installation is available at https://youtu.be/pUcrYwbNQ5Y
[5] A video of the Tampere performance is available at https://youtu.be/16PshXGcrjM
[6] A video of the Strings On-Line installation is available at https://youtu.be/h_6M-_ZPYpA

Figure 4.1: The autonomous "players" of the Strings On-Line installation: a collective of self-playing guitars and Dr Squiggles robots (Photo: Michael Krzyzaniak)

4.3.2 Selected Performances

INTIMAL

I developed the audio program for the sonification of performer's breathing signals. The networked performance took place in Oslo, Barcelona, and London simultaneously (Diaz et al., 2019), during which I also operated the transmission of the sensor data over the network.[7]

No Musicians' Land

In 2020, during the first wave of the coronavirus pandemic, I was invited to play two sets of solo performances using RAW (Section 3.3). The performance were live-streamed in Istanbul, Turkey, as part of an exhibition called *Flux*, organized through the collaboration of Marina Abramović, the Marina Abramović Institute (MAI) and Sakıp Sabancı Museum (SSM).[8]

Fibres Out of Line

I used the musical AI system presented in Paper V, CAVI, also in an interactive art installation and a performance for the 2021 Rhythm Perception and Production Workshop (RPPW). In the performance, a dancer interacted through the network with a number of autonomous musical agents, including rhythm-playing robots, organ-playing robots, and CAVI (see Figure 4.2 for the setup). Visitors could

[7] An audio recording of the INTIMAL performance is available at https://youtu.be/m30yRwG1Tp8

[8] A video excerpt of the live-streamed performance for the No Musicians' Land exhibition is available at https://youtu.be/ikan7NbPTAM

Figure 4.2: The lab environment where musical robots and CAVI were set up for the installation and performance as part of *Fibres Out of Line*.

watch the performance, and subsequently interact with the installation, all remotely via Zoom.[9]

4.3.3 Releases

Bahçe

Bahçe is a duo improvisation album that was released in the Fall of 2020. The duo consisted of a classical guitarist, Yurdal Çağlar, and myself performing on no-input mixer (Section 2.1) together with custom built electronics and effects devices.[10]

[9] A video of the Fibres Out of Line performance is available at https://youtu.be/Txra_hp-H4g

[10] *Bahçe* is available in digital platforms for streaming, such as Spotify at https://open.spotify.com/album/3hJCAK7m5OH7oRgJ6rh3IP?si=32h1qUSdQSaIUcivEKO_hg

Chapter 5

Discussion

The machine is made by humans;
it cannot live without humans.
– A special friend

5.1 Summary

The main objective of this dissertation was to explore shared control between human performers and artificial agents in interactive performance to expand our understanding of agency and musical AI. In this section, I will reflect on the research questions posed in Chapter 1 and discuss some topics that have emerged during the research.

5.1.1 How to include embodied perspective in developing musical agents for interactive performance?

To fulfill the main objective of this dissertation, I asked an overarching *How?* question that I explored via four main projects. These projects were structured according to three operational levels of body movement, from motion to gesture, which I clarified in Section 2.3.4. Accordingly, the first question (RQ1) concerned the physical motion and sound signals found in guitar performance. To answer that, I conducted laboratory experiments to collect datasets and conduct analyses. My second question (RQ2) aimed to explore the use of various artificial intelligence (AI) techniques and methods for embodied interaction with sound- and music-making machines. In doing so, I developed several interactive music systems and evaluated them in public performances. The third question (RW3) was more open-ended than the first two and was interested in the higher-level aspects of performing music with machines. The specific high-level concept I focused on was the agency, and more concretely, the meaning of agency in interactive contexts. This I explored through a public event for one of the interactive systems I developed and an online study we conducted separately. In the following, I will touch upon the prominent findings regarding each of my research questions.

5.1.2 RQ1: What are the relationships between action and sound in instrumental performance, and how can such relationships be used to create new interactive paradigms?

Actions are temporal chunks of the continuous motion signal. We interpret them by subjectively determining their start and end points. In action–sound data collected from thirty-three guitarists as part of the project presented in Paper IV, we made the segmentation based on metronome timestamps. Controlled conditions, such as metronome beats, predefined form, fixed environment, and equipment, facilitated a relatively objective analysis of actions. Still, at the micro-interaction level, each player's actions are unique. A similar segmentation of free improvisation data would be hardly possible, if at all.

The relationships between action and sound demonstrate more significant similarities in the excitation action compared to the modification. The EMG patterns and their resemblance to the resultant sound in sound-producing actions are dependent on the task's difficulty level. We found in the data that as the tasks become more challenging, such as agility, the patterns of the exerted effort become more peculiar. In addition, we observed that participants tend to unwittingly add ornaments, such as vibrato, while sustaining the sound. That makes the modification action unique to the player, resulting in a weaker correlation with the resultant sound dynamics.

For the second part of the question, our strategy was to train a long short-term memory (LSTM) network to map the EMG signals with the energy parameter of the sound synthesis. Our dataset consisted of tasks and free improvisations. Drawing on the embodied music cognition conceptions that suggest such functional categories, our premise was that all human motion could be seen as co-articulations of three basic motion types (impulsive, sustained, and iterative). Since deep learning techniques can often be computationally expensive, we trained nine different model configurations to test the prediction accuracy versus latency. The results were satisfactory because all model configurations were able to predict the sound energy envelope of free improvisations based on a training dataset of solely basic action types. However, even the most miniature model was subject to a perceptible latency.

5.1.3 RQ2: What can AI offer for the action capabilities in interactive systems?

The action capability in an interactive music system refers to the range of actions that can be performed. Consider the example of a simple instrument with a force-sensing resistor (FSR) mapped to a specific sound frequency. Your action capability on that instrument is bound to change the pitch by pressing. Then, add an *if…then* statement in the code so that every time a threshold is met, a random number is generated automatically and adjusts a synthesizer parameter, such as the frequency of a low-frequency oscillator (LFO). You may like the outcome or not, but now the instrument affords a new action capability. Your

goals, hence your actions on the device, will most likely be influenced by such a simple agent.

Across the six papers included in this dissertation, four of them focused on developing new interactive systems that explored the use of musical agents and their influence on the performer's action capability. Paper I presented a new instrument, of which the control was shared between a dancer and a musician. The dancer, through the wearable sensors and a microphone, moves in the air to produce sound. While doing so, the musician controls the sound parameters and interferes with the dancer's action–sound mappings by, for example, changing the scaling. The dancer described her experience of performing with the system as a *new physical language*. From the perspective of the musician, the dancer became an autonomous agent that he could steer. The agent's autonomy enacted a new range of performability, regardless of the musician's lack of control.

In the "air instrument" presented in Paper III, several algorithmic approaches were implemented to afford a different dimension of controllability. The system's control interface was based on two Myo armbands, each worn on a forearm. Thus, an independent finger control, as most acoustic instruments offer, was missing. To tackle that, the system employed a classifier that recognizes the performer's pinch grip to trigger a new musical event. Such an event was a melodic trajectory around the orbit of a chosen strange attractor. This way, the performer was able to create melodic lines. In the other control structure of the instrument, which I described as a gamification strategy, the performer was assigned to intersect two imaginary balls in the air, which corresponded to two points on an XY plane, to trigger a new musical event. In addition, the instrument was automating sonic and musical parameters via real-time audio signal analysis from the rest of the ensemble.

Paper IV explored if we can *translate* the embodied knowledge of an existing instrument into a new interaction paradigm. The results of the modeling approach showed the potential of using machine learning on recorded performance data. Details of this project can be found above as part of the answer to the first question.

The deep learning framework in the project for the action–sound mappings was subject to a considerable latency. According to Boden & Edmonds (2009), there is a difference between *interaction*, leading to a solid action–sound causality, and *influence*, which has rather long-term effects on the output (latency). Drawing on that, we shifted our approach from discriminative to generative modeling in the project of Paper V. The coadaptive audiovisual instrument CAVI generates its own control signals based on muscle activation (EMG) and acceleration (ACC) signals from the performer. The generated control signals automate the live sound processing based on layered time-based effects modules. This setup drastically influenced the range of actions of the performers, a thorough evaluation of which is the subject of the next question.

Every implementation of a new action–sound mapping algorithm denotes a new action capability, which comes with a new set of constraints. AI can add a dynamic aspect that can adapt, alter, surprise, enrich, or compete with the performer's music-making goals.

5.1.4 RQ3: What is the meaning of agency in interactive contexts?

There are two main dimensions of agency. The first is concerned with the term's definition and and an entity's necessary properties to be considered an agent. In Section 2.2.6, I presented a brief review of this dimension, which mostly covered perspectives from computational and cognitive theories. According to these perspectives, an agent can be as simple as an *if...else* condition (Russell, 2010; Schlosser, 2019). As for the agency in interactive contexts, Dahlstedt (2021, p. 31) depicts it clearly in a "spectrum of agency" diagram. What he calls *influential agency* originates in the tool designer and maker and flows onward through the tool's own mediation, the tool user, artwork itself, and, finally, the spectator. Such a flow that yields the artistic result, Dahlstedt maintains, can also be subject to circular detours. These include the art history if the practice is akin to some historical context, personal histories of the persons involved (e.g., makers, performers, etc.), and the aesthetic context.

Two aspects of agency are still under-explored. First, the sense of agency (SoA), meaning the sense of control over the consequences of one's actions. Second, what properties of agents in the particular context of interactive arts that people tend to attribute the agency. Following a brief review of theories presented in Section 2.3.5, I investigated the first aspect, using CAVI in a public event with two improvisers (Paper V). Drawing on my own observations and one-to-one interviews with the performers, the SoA, hence whether the interactive system is experienced as a tool or an actor, varies throughout a performance. When the prediction of the internal comparator mechanism (Wolpert et al., 1995) is violated, e.g., when there is a latency, the SoA gets weaker. However, perceived qualities of the collaborative performance can enhance the SoA. This varies from one musician to another or from whether the performed piece is a composition or a free improvisation.

Throughout my research, the surprise concept emerged as an essential element in the experience of agency. The ambiguity of surprise can be favorable in aesthetic experiences. Therefore, it can build up the perceived qualities that compensate for the prediction errors stemming from joint actions or the lack of causality. In the study I contributed to Paper VI, we found that visitors ascribed the most agency when the mapping between their mouse cursors and the widget (animated eyes) was subject to an angular offset drifting over time using Brownian motion. Finally, both the sensorimotor and perceptual cues of SoA depend highly on environmental factors, such as the physical space and technical setup.

In sum, similar to the concept of gesture as presented in Section 2.3.4, the "meaning" of agency in interactive contexts is a high-level cognitive concept that always involves someone (or something) executing an action and a perceiver. I also believe that the communicative aspect of agency should be seen as a future research direction of musical AI systems.

5.2 General Discussion

The technologies we use to make music carry agency and how we use these technologies strongly influence the music we make. The cultural, social, and political aspects of music and musical experiences were beyond this dissertation's scope. However, these aspects are still present and should not be disregarded. Current music technology research, of which this dissertation is one example, is part of shaping the future of music, and, I believe, the future of humanity at large. The projects and papers presented in this dissertation are just drops in the sea of possible approaches. Still they contribute to an understanding of AI through the lens of embodied music cognition.

AI tools are ubiquitous these days. These tools can be as simple as a mouse click to generate significant portions of a song. They are often statistically based, thus learning from similarities to generate similar results. Data-driven models might pose optimal solutions to certain problems. For example, using AI, video creators can bypass copyright constraints (Frid et al., 2020), or producers can easily add orchestral arrangements to their tracks using music generation plugins, e.g., the *Orb Composer*.[1] However, we should not forget about the algorithmic biases of big data frameworks (Johnson, 2020; Bogroff & Guegan, 2019). How can we consider the agency carried by these tools independently from their potential biases in terms of musical aesthetics? These are important questions, although I have been more interested in how far AI can go, not only in terms of solving pre-defined problems with the utmost accuracy but also how it can become something other than a mere tool.

In the theoretical discussion of this dissertation, I aimed to sketch a picture of two aspects related to such an overarching goal. First, I touched upon cybernetic artists' process-oriented vision to experiencing art in Section 2.1. This vision originated in a conceptual depart from what Ascott (2002) describes as an object-oriented construct of the art before the 20$^{\text{th}}$ century. That construct focused on the virtuosi and the genius of the composer. In contrast, the process-oriented vision focused on the lived experience bursting as a feedback loop circulating through the triad artist/artwork/observer. Second, in Sections 2.2 and 2.3, I tried to point to some parallels between the evolution of AI and a Western music-theoretical approach to music. The latter has received criticism for being object-oriented. One can find the same inclination in AI since its inception in the 1950s. Even though mainstream AI left the symbolicist approach long ago, I argue that current AI is still predominantly object-oriented in being focused on the result, such as composing or painting in the style of an acclaimed artist. On the one hand, the availability of accuracy measures facilitates the engineering of new tools. On the other, that deepens the ontological gap between what the industry wants to achieve and what art does. According to Dahlstedt (2021, p. 32), the former "aims the middle of the circle," whereas the latter, "to extend" it. A significant motivation behind this dissertation is the curiosity and urge to extend the circle.

[1] https://www.orb-composer.com/

With that urge, I highlighted some critiques directed to AI from several cognitive scientists and philosophers for avoiding the embodied perspective. Thus, I also argue that focusing on objects or results goes hand in hand with the disembodied approach prevalent within AI research. Then descriptive properties, such as an approximation accuracy or user-friendliness become more important than purposiveness of collaboration, meaning and emergence. In the foreword of the recent *Handbook of Artificial Intelligence for Music*, Luc Steels writes (Miranda, 2021, p. xvi):

> I do not believe that the rich web of meanings that we as humans naturally engage in will ever be captured by an AI system, particularly if it is disembodied and has no social role in a human community.

My aim has been to explore how to include embodied perspectives in musical agents. Employing machine learning (ML) algorithms to interact with machines using body movement is not new (see Section 2.2.3). Still, I propose a conceptual shift from using the machine as a tool to using it as an actor in musical interaction. That is how I ended up with what I call *shared control*, inspired by early avant-garde and a process-oriented vision. The ambition was to develop instruments that become actors in performance.

The question of determining the measures of performer experience remains open. That is where the concept of agency, or, more specifically, the sense of agency (SoA) emerged while conducting this dissertation project. The investigation of SoA inevitably prioritized the realization of artistic works and collecting qualitative data of performers. Even though I do not claim any factual finding in that respect, the feedback from performers who took part in the evaluations of the systems I developed shed light on future research directions.

As for now, we roughly categorize music-making machines depending on their autonomous features, complexity, or influential agency. Expanding our understanding of the performers' varying sense of control concerning these features is equally critical. For example, in Section 2.3.5, I referred to several studies suggesting that the agency experience can be highly flexible. Depending on the kind of joint activities or the tools being used, it can alter, extend, or even turn into *landing the self's exclusivity*. That is, precisely, what differentiates a process-oriented perspective from the object-oriented one. Such an understanding will help improve the perceptual monitoring of these systems and enrich the current methods and approaches in the performing arts.

5.3 Implications For Research

The research presented in this dissertation is genuinely interdisciplinary. Many people talk about working across disciplines, but in my experience this is challenging in practice. Different theoretical positions and methodological approaches need to be merged. I have been fortunate to carry out this research project in an environment that nurtures such endeavor. To me, it has been natural to work with multi-method approach, in which the artistic and scientific

perspectives are not merely supporting each other but inseparable. I see this approach as a contribution itself. In addition, there are also some more specific implications for research:

- Muscle-sensing: The coadaptation approach was the leading paradigm following the biofeedback and biocontrol approaches (Section 2.1.4. This dissertation project contributes to this line of research with a particular focus on developing novel algorithmic and performative approaches for the electromyogram (EMG).
- Software: The iterative prototyping processes have resulted in several custom software solutions and machine learning frameworks that are shared openly for others to build on.
- Data sets: As part of the empirical study, a multimodal dataset of EMG, motion capture, audio, and video recordings was collected from a total of thirty-six semi-professional and music student guitar players. Hopefully, this dataset can be analyzed further and used also in other creative projects.
- Music interaction: Drawing on previous work on multi-user instruments, I explored different co-performance and improvisation scenarios between performers from different embodied practices, real-time interaction with ensemble members, and shared control of acoustic musicians and AI.
- Musical AI: Several AI techniques were ecologically evaluated and reported in respective publications. Through a literature review and developed interactive systems, I have proposed some future directions for musical AI research.
- Artistic research: This dissertation conducted basic artistic research that can impact future music making. This research also resulted in a number of creative works performed and exhibited in public events, as well as a music album was released.
- Theory: Various results coming out of this dissertation have provided additional empirical evidence for embodied music cognition concepts. Also, the qualitative feedback of the performers in one of the studies presented in Paper V was compatible with perceptual accounts on the agency experience.
- Literature: The extensive review presented in Chapter 2 combined concepts and theories from multiple disciplines, which provided a background for future research in musical human–computer interaction.

5.4 Future Research

Some people see AI as a threat to human craft and values. Should we be afraid of the future of AI? I do not think we should. Machines may be fascinating artifacts

but they are also quite dumb. I reckon the disembodiment and disconnectedness of AI can be an important factor for such dystopic tendencies. In Section 2.3.5, I referred to Wegner (2002, p. 221) who discussed how people can project action to imaginary agents, such as supernatural beings, which he describes as *virtual agency*. We collaborate with AI every day, in one way or another. But we do not actively perceive it regardless of the magnitude of its influence on our actions. Diversifying the artistic repertoire is therefore crucial for the spectator to stop being a passive receptor and experience different control structures involving both humans and machines.

In one of the interviews I conducted as part of this dissertation, a musician mentioned the iconic movie, *The Terminator*, when thinking about AI. The film depicts a war between humans and machines. In the movie, the Terminator brags about himself and say: "My CPU is a neural-net processor; a learning computer. The more contact I have with humans, the more I learn." There was no mention of a body. How would he even be able to contact humans without a body? In this dissertation, I have tried to emphasize the importance of embodiment in musical AI. I will conclude by pointing to three main research directions that I believe are crucial for the further advancement of musical AI:

Embodiment: In the projects presented as part of this dissertation, I primarily focused on the perceptual monitoring systems of the artificial agents. A significant portion of the works in musical AI and multi-agent systems (MAS) focused on the auditory modality by developing fascinating systems that used various methods to track, analyze, and generate sound both in the symbolic and audio domains. Since one of my core arguments is that music is an embodied experience, I put effort into the aspects of the human body that machines can interact with. However, that is only a part of the embodiment, a multimodal construction based on a tight coupling between perception and action. Therefore, I believe that embodiment, physical or virtual, is crucial for AI in general and musical AI in particular. Only in my last project (Papers V and VI), I experimented with a (virtual) embodiment of the musical agent. I will continue to explore such embodiment in the future.

Communication: While focusing on agents' perceptual monitoring of the human body, I drew on the functional categories of music-related movement presented in Section 2.3.4 and Paper II. I discussed these movement categories in terms of a three-level hierarchy. In the projects that formed this dissertation, I experimented with low-level physical motion signals and mid-level actions as goal-directed cognitive chunks. However, a fundamental aspect of collaborative music performance is sociability, hence communication. Then the gestures that denote high-level meaning are indispensable components of communication. Embodied communication is ordinary for humans, but poses great challenges for computers. Significant work has been done in music research over the last years (see, e.g., Schiavio & Høffding (2015); Bishop et al. (2019); Bishop & Goebl (2020)) investigating the embodied communication between performers. In the future, I aim to build on a similar approach in developing cognitive musical agents with hierarchical architectures. This includes research into the meaning-related high-level aspects of body movement, such as gestural expressions of

affect, memory and intentions.

Self-evaluation: The concept of feedback is vital for the self-regulation of both living organisms and artificial agents. In the particular context of musical AI and MAS, such mechanisms range from low-level techniques (see, e.g., Holopainen (2012)) to higher-level cognitive measures, e.g., sound affect estimation (Russell, 1980). My literature review revealed a gap in that most methods and techniques used are either monomodal or do not account for the concept of control. The particular study we presented in Paper VI showed that interactive engagement is related to the amount of agency attributed to artificial agents. Can new patterns emerge in an interactive system just through spontaneous negotiation with human performers without having any predefined shared control structure? That is where the idea of SoA gains importance. SoA models have been proposed in the field of AI (see, e.g., (Legaspi et al., 2019)). In the future, I aim to explore these approaches within reward mechanisms of agents and investigate the measures for automating the tracking of human performers' varying agency experiences.

Bibliography

Ahissar, E. & Assa, E. (2016). Perception as a closed-loop convergence process. *eLife*, *5*, e12830. Publisher: eLife Sciences Publications, Ltd.

Anchor, K. N., Beck, S. E., Sieveking, N. & Adkins, J. (1982). A history of clinical biofeedback. *American Journal of Clinical Biofeedback*, *5(1)*, 3–16. Place: Canada Publisher: Hans Huber Publishers.

Apollonius, A. a. (250BC). Kitab Arshimidas fi al-binkamat and San'at al-zamir.

AQAXA (2021). Corporeal EP. https://aqaxa.bandcamp.com/album/corporeal-ep-2.

Ascott, R. (1968). The Cybernetic Stance: My Process and Purpose. *Leonardo*, *1(2)*, 105–112. Publisher: The MIT Press.

Ascott, R. (2002). Behaviourist Art and the Cybernetic Vision. In R. Packer & K. Jordan (Eds.), *Multimedia. From Wagner to Virtual Reality* (pp. 104–120). New York, London: W. W. Norton & Company.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. Springer US.

Bailey, D. (1993). *Improvisation: its nature and practice in music*. New York: Da Capo Press.

Baldan, S., Delle Monache, S. & Rocchesso, D. (2017). The Sound Design Toolkit. *SoftwareX*, *6*, 255–260.

Benson, C., Manaris, B., Stoudenmier, S. & Ward, T. (2016). SoundMorpheus: A Myoelectric-Sensor Based Interface for Sound Spatialization and Shaping. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 332–337). Brisbane, Australia: Zenodo.

Berdahl, E., Sheffield, E., Pfalz, A. & Marasco, A. T. (2018). Widening the Razor-Thin Edge of Chaos Into a Musical Highway: Connecting Chaotic Maps to Digital Waveguides. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 390–393). Blacksburg, Virginia, USA: Zenodo.

Berry, R. & Dahlstedt, P. (2003). Artificial Life: Why Should Musicians Bother? *Contemporary Music Review*, *22(3)*, 57–67. Publisher: Routledge _eprint: https://doi.org/10.1080/0749446032000150889.

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F. & Rasamimanana, N. (2010). Continuous Realtime Gesture Following and Recognition. In Kopp, S. & Wachsmuth, I. (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction*, Lecture Notes in Computer Science (pp. 73–84). Berlin, Heidelberg: Springer.

Beyls, P. (2007). Interaction and Self-organisation in a Society of Musical Agents. In *Proceedings of MusicAL: Workshop on Music and Artificial Life*.

Bishop, C. M. (1994). Mixture density networks. Technical Report NCRG/97/004, Neural Computing Research Group, Aston University„ Birmingham. Publisher: Aston University.

Bishop, L., Cancino-Chacón, C. & Goebl, W. (2019). Moving to Communicate, Moving to Interact. *Music Perception*, *37(1)*, 1–25.

Bishop, L. & Goebl, W. (2020). Negotiating a Shared Interpretation During Piano Duo Performance. *Music & Science*, *3*, 2059204319896152. Publisher: SAGE Publications Ltd.

Boden, M. A. (1977). *Artificial intelligence and natural man*. New York: Harvester Press.

Boden, M. A. (1996). *The Philosophy of Artificial Life*. Oxford University Press.

Boden, M. A. (2006). *Mind as machine: a history of cognitive science*. Oxford, New York: Clarendon Press ; Oxford University Press.

Boden, M. A. (2015). Creativity and ALife. *Artificial Life*, *21(3)*, 354–365.

Boden, M. A. & Edmonds, E. A. (2009). What is generative art? *Digital Creativity*, *20(1-2)*, 21–46. Routledge.

Bogroff, A. & Guegan, D. (2019). Artificial Intelligence, Data, Ethics An Holistic Approach for Risks and Regulation. *SSRN Electronic Journal*.

Borgo, D. (2002). Negotiating Freedom: Values and Practices in Contemporary Improvised Music. *Black Music Research Journal*, *22(2)*, 165–188.

Borgo, D. (2005). Rivers of Consciousness: The Nonlinear Dynamics of Free Jazz. In *Jazz Research Proceedings Yearbook* (pp. 46–58).

Borgo, D. & Kaiser, J. (2010). Configurin(g) KaiBorg: Interactivity, ideology, and agency in electro-acoustic improvised music.

Bratman, M. E. (1992). Shared Cooperative Activity. *The Philosophical Review*, *101(2)*, 327–341. Duke University Press.

Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H. H. O. & Philipsen, A. (2018). The Senses of Agency and Ownership: A Review. *Frontiers in Psychology*, *9*. Publisher: Frontiers.

Bretan, M., Gopinath, D., Mullins, P. & Weinberg, G. (2016). A Robotic Prosthesis for an Amputee Drummer. *arXiv:1612.04391*.

Briot, J.-P. & Pachet, F. (2018). Music Generation by Deep Learning - Challenges and Directions. *Neural Computing and Applications*. arXiv: 1712.04371.

Brooks, R. A. (1991a). Intelligence Without Reason. In Mylopoulos, J. & Reiter, R. (Eds.), *IJCAI'91: Proceedings of the 12th International Joint Conference on Artificial Intelligence - Volume 1* (p.2̃8). Sydney New South Wales Australia: Morgan Kaufmann Publishers Inc. The International Joint Conferences on Artificial Intelligence, Inc.

Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence*, *47(1)*, 139–159.

Buchner, A. (1978). *Mechanical Musical Instruments*. Westport, CT, US: Greenwood Press.

Buckner, C. & Garson, J. (2019). Connectionism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 Ed.). Metaphysics Research Lab, Stanford University.

Bullock, J. & Momeni, A. (2015). ml.lib: Robust, Cross-platform, Open-source Machine Learning for Max and Pure Data. In *Proceedings of the international conference on New Interfaces for Musical Expression* (pp. 265–270). Baton Rouge, Louisiana, USA: Louisiana State University.

Burgin, M. (2017). Systems, Actors and Agents: Operation in a multicomponent environment. *arXiv:1711.08319*. arXiv: 1711.08319.

Burns, C. (2002). Realizing Lucier and Stockhausen: Case Studies in the Performance Practice of Electroacoustic Music. *Journal of New Music Research*, *31(1)*, 59–68. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1076/jnmr.31.1.59.8104.

Cadoz, C. (1988). Instrumental Gesture and Musical Composition. In *ICMC 1988 - International Computer Music Conference* (pp. 1–12). Cologne, Germany.

Cadoz, C. & Wanderley, M. M. (2000). Gesture - Music. In I.-C. P. Marcelo Wanderley et Marc Battier (Ed.), *Trends in Gestural Control of Music*.

Cage, J. (1961). *Silence: lectures and writings*. Wesleyan University Press. Hanover, NH.

Cage, J. (1991). An Autobiographical Statement. *Southwest Review*, *76(1)*, 59–76. Publisher: Southern Methodist University.

Cage, J. & Goldberg, J. (1976). John Cage: Interviewed by Jeff Goldberg. *The Transatlantic Review*, *(55/56)*, 103–110. Publisher: Joseph F. McCrindle Foundation.

Camurri, A. (1993). Applications of Artificial Intelligence Methodologies and Tools for Music Description and Processing. In G. Haus (Ed.), *Music Processing, The Computer Music and Digital Audio Series* (pp. 233–266). A-R Editions.

Camurri, A. & Coglio, A. (1998). An architecture for emotional agents. *IEEE MultiMedia*, *5(4)*, 24–33. Conference Name: IEEE MultiMedia.

Camurri, A. & Leman, M. (1997). AI-based Music Signal Applications - A Hybrid Approach. In C. Roads, S. Pope, A. Piccialli & G. De Poli (Eds.), *Music Signal Processing* (pp. 349–381). Swets & Zeitlinge.

Cantrell, M. (2007). Enactive Reading: John Cage, Chance, and Poethical Experience. *Genre*, *40(1-2)*, 131–156.

Caramiaux, B., Bevilacqua, F. & Tanaka, A. (2013). Beyond recognition: using gesture variation for continuous interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (p. 2109). Paris, France: ACM Press.

Caramiaux, B. & Donnarumma, M. (2021). Artificial Intelligence in Music and Performance: A Subjective Art-Research Inquiry. In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (pp. 75–95). Cham: Springer International Publishing.

Caramiaux, B., Donnarumma, M. & Tanaka, A. (2015). Understanding Gesture Expressivity through Muscle Sensing. *ACM Transactions on Computer-Human Interaction*, *21(6)*, 1–26.

Caramiaux, B., Françoise, J., Schnell, N. & Bevilacqua, F. (2014a). Mapping Through Listening. *Computer Music Journal*, *38(3)*, 34–48.

Caramiaux, B., Montecchio, N., Tanaka, A. & Bevilacqua, F. (2014b). Adaptive Gesture Recognition with Variation Estimation for Interactive Systems. *ACM Transactions on Interactive Intelligent Systems*, *4(4)*, 1–34.

Caramiaux, B., Wanderley, M. M. & Bevilacqua, F. (2012). Segmenting and Parsing Instrumentalists' Gestures. *Journal of New Music Research*, *41(1)*, 13–29.

Carr, C. J. & Zukowski, Z. (2018). Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands. *arXiv:1811.06633*.

Chabot, X., Dannenberg, R. & Bloch, G. (1986). A Workstation in Live Performance: Composed Improvisation. In *Proceedings of the International Computer Music Conference*.

Chadabe, J. (2002). The Limitations of Mapping As a Structural Descriptive in Electronic Instruments. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression* (pp. 1–5). Dublin, Ireland: Media Lab Europe.

Charrieras, D. & Hochherz, O. (2016). Chasing after the mixer. In *R>EVOLUTION. Proceedings of the 22nd International Symposium on Electronic Arts* (pp. 253–254).

Chi, D., Costa, M., Zhao, L. & Badler, N. (2000). The EMOTE model for effort and shape. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00 (pp. 173–182). USA: ACM Press/Addison-Wesley Publishing Co.

Churchland, A. K. (2011). Normalizing relations between the senses. *Nature Neuroscience*, *14(6)*, 672–673.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3(9)*, 345–351.

Clark, A. (2004). Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence.

Clark, A. (2008). The Negotiable Body. In *Supersizing the Mind*. New York: Oxford University Press.

Clayton, M. (2012). What is Entrainment? Definition and applications in musical research. *Empirical Musicology Review*, *7(1-2)*, 49–56.

Collins, N. & Lonergan, S. (Eds.). (2020). *Handmade Electronic Music: The Art of Hardware Hacking* (3 Ed.). New York: Routledge.

Collins, N., Schedel, M. & Wilson, S. (2013). The post-war sonic boom. In *Electronic Music*, Cambridge Introductions to Music (pp. 45–64). Cambridge: Cambridge University Press.

Collins, N. M. (2006). *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems*. PhD thesis, University of Cambridge.

Cont, A. (2010). A Coupled Duration-Focused Architecture for Real-Time Music-to-Score Alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 974–987. Publisher: Institute of Electrical and Electronics Engineers.

Cope, D. (1989). Experiments in musical intelligence (EMI): Non-linear linguistic-based composition. *Interface*, *18(1-2)*, 117–139. Routledge.

Cope, D. (2001). *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, MA, USA: MIT Press.

Copeland, B. J. & Long, J. (2017). Turing and the History of Computer Music. In J. Floyd & A. Bokulich (Eds.), *Philosophical Explorations of the Legacy of Alan Turing: Turing 100*, Boston Studies in the Philosophy and History of Science (pp. 189–218). Cham: Springer International Publishing.

Crispin, D. & Gilmore, B. (2014). *Artistic experimentation in music: an anthology*. Leuven: Leuven University Press.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper & Row.

Côté-Allard, U., Fall, C. L., Drouin, A., Campeau-Lecours, A., Gosselin, C., Glette, K., Laviolette, F. & Gosselin, B. (2019). Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27(4)*, 760–771.

Dahlstedt, P. (2007). Autonomous Evolution of Complete Piano Pieces and Performances.

Dahlstedt, P. (2018). Action and Perception: Embodying algorithms and the extended mind. In R. T. Dean & A. McLean (Eds.), *The Oxford Handbook of Algorithmic Music*, Volume 1. Oxford University Press.

Dahlstedt, P. (2019). Big Data and Creativity. *European Review*, *27(3)*, 411–439. Publisher: Cambridge University Press.

Dahlstedt, P. (2021). Musicking with Algorithms: Thoughts on Artificial Intelligence, Creativity, and Agency. In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (pp. 873–914). Cham: Springer International Publishing.

Di Donato, B., Bullock, J. & Tanaka, A. (2018). Myo Mapper: A Myo Armband To Osc Mapper. *Zenodo*.

Di Donato, B. & Dooley, J. (2017). MyoSpat: a system for manipulating sound and light projections through hand gestures. Salford, Manchester.

Diaz, X. A., Sanchez, V. E. G. & Erdem, C. (2019). INTIMAL: Walking to Find Place, Breathing to Feel Presence. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 246–249). Porto Alegre, Brazil: Zenodo.

Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M. & Saurous, R. A. (2017). TensorFlow Distributions. *arXiv:1711.10604*.

Dixon, S. (2017). Cybernetic-existentialism in interactive performance: strangers, being-for-others and autopoiesis. *International Journal of Performance Arts and Digital Media*, *13(1)*, 55–76. Publisher: Routledge _eprint: https://doi.org/10.1080/14794713.2017.1301173.

Dixon, S. (2019). Cybernetic-Existentialism in Performance Art. *Leonardo*, *52(3)*, 247–254. Publisher: The MIT Press.

Doel, K. v. d. (2005). Physically based models for liquid sounds. *ACM Transactions on Applied Perception*, *2(4)*, 534–546.

Donahue, C., McAuley, J. & Puckette, M. (2019). Adversarial Audio Synthesis (p.1̃6).

Donnarumma, M. (2011). Xth Sense: researching muscle sounds for an experimental paradigm of musical performance (p.9̃).

Donnarumma, M. (2016). *Configuring Corporeality: Performing bodies, vibrations and new musical instruments*. PhD thesis, Goldsmiths, University of London, London, United Kingdom.

Donnarumma, M. & Pevere, M. (2018). Eingeweide. https://marcodonnarumma.com/works/eingeweide/.

Dourish, P. (2001). *Where the action is: the foundations of embodied interaction*. Cambridge, Mass: MIT Press.

Dreyfus, H. L. (1987). Misrepresenting Human Intelligence. In *Artificial Intelligence*. Routledge.

Dreyfus, H. L. (2001). Phenomenological Description Versus Rational Reconstruction. *Revue Internationale de Philosophie*, *55(216 (2))*, 181–196. Publisher: Revue Internationale de Philosophie.

Dromey, C., Reese, L. & Hopkin, J. A. (2009). Laryngeal-Level Amplitude Modulation in Vibrato. *Journal of voice*, *23(2)*, 156–163. Place: New York Publisher: Mosby, Inc.

Dubnov, S. & Assayag, G. (2005). Improvisation Planning and Jam Session Design Using Concepts of Sequence Variation and Flow Experience. Salerno, Italy: Zenodo.

Dudas, R. (2010). "Comprovisation": The Various Facets of Composed Improvisation within Interactive Performance Systems. *Leonardo Music Journal*, *20(1)*, 29–31.

Eck, D. & Schmidhuber, J. (2002). Finding temporal structure in music: blues improvisation with LSTM recurrent networks. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing* (pp. 747–756).

Ellefsen, K. O., Martin, C. P. & Torresen, J. (2019). How do Mixture Density RNNs Predict the Future? *arXiv:1901.07859*.

Engel, J., Agrawal, K. K., Chen, S., Gulrajani, I., Donahue, C. & Roberts, A. (2018). GANSynth: Adversarial Neural Audio Synthesis. In *Proceedings of the International Conference on Learning Representations*. Vancouver, BC.

Engel, J., Hantrakul, L., Gu, C. & Roberts, A. (2020). DDSP: Differentiable Digital Signal Processing. *arXiv:2001.04643 [cs, eess, stat]*. arXiv: 2001.04643 version: 1.

Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D. & Simonyan, K. (2017). Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 1068–1077). PMLR. ISSN: 2640-3498.

Englehart, K. & Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, *50(7)*, 848–854.

Engramelle, M.-D.-J. (1775). *La Tonotechnie ou l'art de noter les cylindres et tout ce qui est susceptible de notage dans les instrumens de concerts méchaniques... par le père Engramelle,....* A Paris: chez P. M. Delaguette. OCLC: 495086725.

Erdem, C. (2020). Towards Playing in the 'Air'. doi:10.5281/zenodo.6478033.

Erdem, C. (2021). CAVI. doi:10.5281/zenodo.6478027.

Erdem, C., Camci, A. & Forbes, A. (2017). Biostomp: A Biocontrol System for Embodied Performance Using Mechanomyography. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 65–70). Copenhagen, Denmark: Zenodo.

Erdem, C. & Jensenius, A. R. (2020). RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 477–482). Birmingham, UK: Zenodo.

Erdem, C., Lan, Q. & Jensenius, A. R. (2020). Exploring relationships between effort, motion, and sound in new musical instruments. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *16(3)*, 310–347.

Erdem, C., Schia, K. H. & Jensenius, A. R. (2019). Vrengt: A Shared Body-Machine Instrument for Music-Dance Performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 186–191). Porto Alegre, Brazil: Zenodo.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (Revised Edition Ed.). Cambridge, MA, USA: A Bradford Book.

Farmer, H. G. (1931). *The organ of the ancients from Eastern sources, Hebrew, Syriac and Arabic.* London: William Reeves.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J. & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *NeuroImage*, *18(2)*, 324–333.

Farrer, C. & Frith, C. D. (2002). Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage*, *15(3)*, 596–603.

Fdili Alaoui, S., Bevilacqua, F., Bermudez Pascual, B. & Jacquemin, C. (2013). Dance interaction with physical model visuals based on movement qualities. *International Journal of Arts and Technology*, *6(4)*, 357–387.

Fellgett, P. (1988). Cybernetics: Retrospect and Prospect. *Kybernetes*, *17(3)*, 22–31. Publisher: MCB UP Ltd.

Ferguson, J. R. (2013). Imagined Agency: Technology, Unpredictability, and Ambiguity. *Contemporary Music Review*, *32(2-03)*, 135–149. Publisher: Routledge _eprint: https://doi.org/10.1080/07494467.2013.775810.

Fernandez, J. D. & Vico, F. (2013). AI Methods in Algorithmic Composition: A Comprehensive Survey. *Journal of Artificial Intelligence Research*, *48*, 513–582. arXiv: 1402.0585.

Fiebrink, R. & Caramiaux, B. (2016). The Machine Learning Algorithm as Creative Musical Tool. *arXiv:1611.00379 [cs]*. arXiv: 1611.00379.

Fiebrink, R. A. (2011). *Real-time human interaction with supervised learning algorithms for music composition and performance.* phd, Princeton University, USA. AAI3445567 ISBN-13: 9781124491899.

Floridi, L. & Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines*, *14(3)*, 349–379.

Foster, D. & Safari, a. O. M. C. (2019). *Generative deep learning: teaching machines to paint, write, compose, and play.* OCLC: 1099922678.

Fourneret, P. & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36(11)*, 1133–1140.

François, C. (1999). Systemics and cybernetics in a historical perspective. *Systems Research and Behavioral Science*, *16(3)*, 203–219.

Françoise, J. (2015). myo for max. https://github.com/JulesFrancoise/myo-for-max.

Françoise, J., Fdili Alaoui, S., Schiphorst, T. & Bevilacqua, F. (2014). Vocalizing dance movement for interactive sonification of laban effort factors. In *Proceedings of the 2014 conference on Designing interactive systems*, DIS '14 (pp. 1079–1082). New York, NY, USA: Association for Computing Machinery.

Frid, E., Gomes, C. & Jin, Z. (2020). Music Creation by Example. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20 (pp. 1–13). New York, NY, USA: Association for Computing Machinery.

Friedman, K., Smith, O. F. & Poggenpohl, S. H. (2005). *Fluxus and legacy: special issue, part 1.* Providence, R.I.: Rhode Island School of Design, Graphic Design Dept. OCLC: 225215706.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4(1)*, 14–21.

Gallagher, S. (2007). The Natural Philosophy of Agency. *Philosophy Compass*, *2(2)*, 347–357. doi:10.1111/j.1747-9991.2007.00067.x.

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119(2)*, 593–609.

Gentsch, A. & Schutz-Bosbach, S. (2015). Agency and Outcome Prediction. In P. Haggard & B. Eitam (Eds.), *The Sense of Agency* (pp. 217–234). Oxford University Press.

Georgieff, N. & Jeannerod, M. (1998). Beyond Consciousness of External Reality: A "Who" System for Consciousness of Action and Self-Consciousness. *Consciousness and Cognition*, *7(3)*, 465–477.

Gibbons, A. (2011). *Multimodality, Cognition, and Experimental Literature*. New York: Routledge.

Gibbs, Raymond W., J. (2005). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, Mass: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. The ecological approach to visual perception. Boston, MA, US: Houghton, Mifflin and Company. Pages: xiv, 332.

Gillian, N. (2011). A machine learning toolbox for musician computer interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Gillian, N. & Paradiso, J. A. (2014). The Gesture Recognition Toolkit. *Journal of Machine Learning Research*, *15(101)*, 3483–3487.

Godøy, R. I. (2003). Motor-Mimetic Music Cognition. *Leonardo*, *36(4)*, 317–319.

Godøy, R. I. (2006). Gestural-Sonorous Objects: embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound*, *11(2)*, 149–157.

Godøy, R. I. (2009a). Chunking Sound for Musical Analysis. In Ystad, S., Kronland-Martinet, R. & Jensen, K. (Eds.), *Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music*, Lecture Notes in Computer Science (pp. 67–80). Berlin, Heidelberg: Springer.

Godøy, R. I. (2009b). Gestural Affordances of Musical Sound. In *Musical Gestures*. Routledge. Num Pages: 23.

Godøy, R. I. (2013). Thinking Sound and Body-Motion Shapes in Music: Public Peer Review of "Gesture and the Sonic Event in Karnatak Music" by Lara Pearson. *Empirical Musicology Review*, *8(1)*, 15–18. Number: 1.

Godøy, R. I. (2018a). Key-postures, trajectories and sonic shapes. In *Music and Shape*. New York: Oxford University Press.

Godøy, R. I. (2018b). Motor Constraints Shaping Musical Experience. *Music Theory Online*, *24(3)*.

Godøy, R. I. (2018c). Sonic Object Cognition. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology* (pp. 761–777). Berlin, Heidelberg: Springer Berlin Heidelberg. Series Title: Springer Handbooks.

Godøy, R. I. (2022). Understanding musical instants. In M. Doffman, E. Payne & T. Young (Eds.), *The Oxford Handbook of Time in Music*, Oxford Handbooks. Oxford, New York: Oxford University Press.

Godøy, R. I., Haga, E. & Jensenius, A. R. (2006). Playing "Air Instruments": Mimicry of Sound-Producing Gestures by Novices and Experts. In S. Gibet, N. Courty & J.-F. Kamp (Eds.), *Gesture in Human-Computer Interaction and Simulation*, Volume 3881 (pp. 256–267). Berlin, Heidelberg: Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.

Godøy, R. I. & Leman, M. (2010). *Musical Gestures: Sound, Movement, and Meaning*. London: Routledge.

Goffey, A. (2008). Algorithm. In *Software Studies*. The MIT Press.

Golan, A. (2019). The Musical Boat for a Drinking Party. https://aljazaribook.com/en/2019/08/07/the-musical-boat-en/.

Goldman, J. (2012). The Buttons on Pandora's Box: David Tudor and the Bandoneon. *American Music*, *30(1)*, 30–60. Publisher: University of Illinois Press.

Gonzalez-Sanchez, V., Dahl, S., Hatfield, J. L. & Godøy, R. I. (2019). Characterizing Movement Fluency in Musical Performance: Toward a Generic Measure for Technology Enhanced Learning. *Frontiers in Psychology*, *10*.

Gonzalez-Sanchez, V. E., Zelechowska, A. & Jensenius, A. R. (2018). Correspondences Between Music and Involuntary Human Micromotion During Standstill. *Frontiers in Psychology*, *9*, 1382.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. Adaptive Computation and Machine Learning series. Cambridge, MA, USA: MIT Press.

Graves, A. (2013). Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850 [cs]*. arXiv: 1308.0850.

Gritten, A. & King, E. (2006). *Music and gesture*. Aldershot: Ashgate.

Gritten, A. & King, E. (Eds.). (2011). *New perspectives on music and gesture*. (SEMPRE Studies in The Psychology of Music). Farnham: Ashgate. Routledge.

Gurevich, M. (2014). Distributed Control in a Mechatronic Musical Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 487–490). London, United Kingdom: Zenodo.

Gurevich, M., Stapleton, P. & Marquez-Borbon, A. (2010). Style and Constraint in Electronic Musical Instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 106–111). Sydney, Australia.

Ha, D. & Eck, D. (2017). A Neural Representation of Sketch Drawings. *arXiv:1704.03477*.

Hadjeres, G., Pachet, F. & Nielsen, F. (2017). DeepBach: a Steerable Model for Bach Chorales Generation. *arXiv:1612.01010 [cs]*. arXiv: 1612.01010.

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences*, *9(6)*, 290–295.

Haggard, P., Clark, S. & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5(4)*, 382–385. Nature Publishing Group.

Haggard, P. & Eitam, B. (Eds.). (2015). *The Sense of Agency*. Social Cognition and Social Neuroscience. New York: Oxford University Press.

Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E. & Alon, U. (2014). Individuality and Togetherness in Joint Improvised Motion. *PLOS ONE*, *9(2)*, e87213. Public Library of Science.

Haskins, R. (2014). Aspects of Zen Buddhism as an Analytical Context for John Cage's Chance Music. *Contemporary Music Review*, *33(5-6)*, 616–629. Publisher: Routledge _eprint: https://doi.org/10.1080/07494467.2014.998426.

Haviv, D., Rivkind, A. & Barak, O. (2019). Understanding and Controlling Memory in Recurrent Neural Networks. *arXiv:1902.07275 [cs, stat]*. arXiv: 1902.07275.

Hayles, N. K. (1999). *How we became posthuman: virtual bodies in cybernetics, literature, and informatics*. Chicago, Ill.: University of Chicago Press. OCLC: 659559883.

Hegarty, P. (2007). *Noise/music: a history*. New York: Continuum. OCLC: 145379732.

Helm, B. (2000). Emotional Reason: How to Deliberate about Value. *American Philosophical Quarterly*, *37(1)*, 1–22. North American Philosophical Publications, University of Illinois Press.

Hermann, T. & Hunt, A. (2011). Interactive Sonification. In T. Hermann, A. Hunt & J. G. Neuhoff (Eds.), *The sonification handbook* (pp. 273–298). Berlin: Logos Verlag. OCLC: ocn771999159.

Hewitt, C. (1976). Viewing Control Structures as Patterns of Passing Messages. Accepted: 2004-10-04T14:48:11Z.

Hewitt, C., Bishop, P. & Steiger, R. (1973). A universal modular ACTOR formalism for artificial intelligence. In *Proceedings of the 3rd international joint conference on Artificial intelligence*, IJCAI'73 (pp. 235–245). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Hickok, G., Buchsbaum, B., Humphries, C. & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience*, *15(5)*, 673–682.

Hill, P. (1974). *The Book of Knowledge of Ingenious Mechanical Devices: (Kitāb fī ma 'rifat al-iyal al-handasiyya)*. Springer Netherlands.

Hiller, L. & Kumra, R. (1979). Composing Algorithms II by means of change-ringing. *Interface*, *8(3)*, 129–168. Publisher: Routledge _eprint: https://doi.org/10.1080/09298217908570271.

Hiller, L. A. & Isaacson, L. M. (1979). *Experimental Music; Composition with an Electronic Computer*. USA: Greenwood Publishing Group Inc.

Hobbes, T. (1929). *Hobbes's Leviathan (Reprint ed, 1651)*. Clarendon Press.

Hobbes, T. (2001). *Of man: being the first part of Leviathan*. New York: Bartleby.com. OCLC: 55992833.

Hochreiter, S. & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9(8)*, 1735–1780.

Hoffman, G. & Weinberg, G. (2011). Interactive Improvisation with a Robotic Marimba Player. In J. Solis & K. Ng (Eds.), *Musical Robots and Interactive Multimodal Systems*, Springer Tracts in Advanced Robotics (pp. 233–251). Berlin, Heidelberg: Springer.

Hoggett, R. (2012a). 1810 - Automaton Trumpet Player - Friedrich Kaufmann (German). http://cyberneticzoo.com/robots/1810-automaton-trumpet-player-friedrich-kaufmann-german/.

Hoggett, R. (2012b). 1849 - Flute-Playing Automaton - Innocenzo Manzetti (Italian). http://cyberneticzoo.com/robots/1849-flute-playing-automaton-innocenzo-manzetti-italian/.

Holbrook, J. B. (2013). What is interdisciplinary communication? Reflections on the very idea of disciplinary integration. *Synthese*, *190(11)*, 1865–1879.

Holopainen, R. (2012). *Self-organised Sound with Autonomous Instruments: Aesthetics and experiments*. PhD thesis, University of Oslo.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79(8)*, 2554–2558. Publisher: National Academy of Sciences Section: Research Article.

Hunt, A. & Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound*, *7(2)*, 97–108.

Huron, D. (2019). Musical Aesthetics: Uncertainty and Surprise Enhance Our Enjoyment of Music. *Current Biology*, *29(23)*, R1238–R1240.

Ivachnenko, A. G. (1967). *Cybernetics and forecasting techniques*, Volume 8 of *Modern analytic and computational methods in science and mathematics*. New York: Elsevier.

Ivakhnenko, A. G., Lapa, V. G., United States & Joint Publications Research Service (1965). *Cybernetic predicting devices,*. New York: CCM Information Corp.

Jeannerod, M. (2008). The sense of agency and its disturbances in schizophrenia: a reappraisal. *Experimental Brain Research*, *192(3)*, 527.

Jeannerod, M. & Pacherie, E. (2004). Agency, Simulation and Self-identification. *Mind & Language*, *19(2)*, 113–146. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0017.2004.00251.x.

Jennings, N. R., Sycara, K. & Wooldridge, M. (1998). A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, *1(1)*, 7–38.

Jensen, M. G. (2009). John Cage, Chance Operations, and the Chaos Game: Cage and the "I Ching". *The Musical Times*, *150(1907)*, 97–102. Publisher: Musical Times Publications Ltd.

Jensenius, A. R. (2007). Action-sound : developing methods and tools to study music-related body movement.

Jensenius, A. R. (2014). To gesture or Not? An Analysis of Terminology in NIME Proceedings 2001-2013. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. London, United Kingdom: Zenodo.

Jensenius, A. R. (2017). Sonic Microinteraction in “the Air”. In M. Lesaffre, P.-J. Maes & M. Leman (Eds.), *The Routledge Companion to Embodied Music Interaction* (1 Ed.) (pp. 429–437). New York ; London : Routledge, 2017.: Routledge.

Jensenius, A. R. (2018a). Methods for Studying Music-Related Body Motion. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, Springer Handbooks (pp. 805–818). Berlin, Heidelberg: Springer.

Jensenius, A. R. (2018b). The Musical Gestures Toolbox for Matlab. In *Late-Breaking/Demo Session Abstracts for the 2018 International Society for Music Information Retrieval Conference* (p.2̃).

Jensenius, A. R. & Lyons, M. J. (Eds.). (2017). *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*. Current Research in Systematic Musicology. Springer International Publishing.

Jensenius, A. R., Sanchez, V. G., Zelechowska, A. & Bjerkestrand, K. A. V. (2017). Exploring the Myo controller for sonic microinteraction. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 442–445). Copenhagen, Denmark: Zenodo.

Jensenius, A. R., Wanderley, M. M., Godøy, R. I. & Leman, M. (2010). Musical Gestures: concepts and methods in research. *978-0-415-99887-1* (pp. 12–35).

Johnson, D. D., Keller, R. M. & Weintraut, N. (2017). Learning to Create Jazz Melodies Using a Product of Experts.

Johnson, G. M. (2020). Algorithmic bias: on the implicit biases of social technology. *Synthese*.

Jordà Puig, S. (2005). *Digital Lutherie – Crafting musical computers for new musics performance and improvisation*. Ph.D. Thesis, Universitat Pompeu Fabra.

Joseph, R. D. (1961). *Contributtiions to Perceptron Theory*. PhD thesis, Cornell University.

Kamkar, S. (2014). Myo-OSC. https://github.com/samyk/myo-osc.

Kelkar, T. (2019). *Computational Analysis of Melodic Contour and Body Movement*. PhD thesis, University of Oslo. Accepted: 2019-11-28T07:48:40Z Publisher: Oslo 07-Media.

Kemper, S. & Cypess, R. (2019). Can Musical Machines Be Expressive? Views from the Enlightenment and Today. *Leonardo*, *52(5)*, 448–454. Publisher: The MIT Press.

Khan, M. A. R. & Poskitt, D. S. (2013). A Note on Window Length Selection in Singular Spectrum Analysis. *Australian & New Zealand Journal of Statistics*, *55(2)*, 87–108. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/anzs.12027.

Kiefer, C. (2014). Musical Instrument Mapping Design with Echo State Networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 293–298). London, United Kingdom: Zenodo.

Kieronski, R. (1966). Vochrome.

King, A. J. (1993). Multisensory Integration: The Merging of the Senses. Barry E. Stein and M. Alex Meredith. MIT Press, Cambridge, MA, 1993. xvi, 211 pp., illus. $42.50 or £38.25. Cognitive Neuroscience Series. *Science*, *261(5123)*, 928–929. Publisher: American Association for the Advancement of Science.

Kingma, D. P. & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*. arXiv: 1412.6980.

Kline, R. R. (2015). *The cybernetics moment: or why we call our age the information age*. OCLC: 890127838.

Knoblich, G., Butterfill, S. & Sebanz, N. (2011). Psychological Research on Joint Action. In *Psychology of Learning and Motivation*, Volume 54 (pp. 59–101). Elsevier.

Kosowitz, S. & Vickery, L. (2013). Retaining a sense of spontaneity in Free Jazz improvisation through music technology. *Research outputs 2013*.

Kotseruba, I. & Tsotsos, J. K. (2020). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53(1)*, 17–94.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, Volume 25. Curran Associates, Inc.

Krzyzaniak, M. (2016). *Timbral Learning for Musical Robots*. PhD thesis, Arizona State University.

Krzyżaniak, M. (2021). Musical robot swarms, timing, and equilibria. *Journal of New Music Research*, *50(3)*, 279–297. Publisher: Routledge _eprint: https://doi.org/10.1080/09298215.2021.1910313.

Lago, N. P. & Kon, F. (2004). The Quest for Low Latency. In *Proceedings of the International Computer Music Conference (icmc2004* (pp. 33–36).

Lahav, A., Saltzman, E. & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27(2)*, 308–314.

Lan, Q. (2019). chaosprint/dual-myo-recorder. https://github.com/chaosprint/dual-myo-recorder.

Latour, B. (2005). *Reassembling the social: an introduction to actor-network-theory*. Oxford University Press.

Lee, M., Freed, A. & Wessel, D. (1991). Real-Time Neural Network Processing of Gestural and Acoustic Signals (pp. 277–280). Montreal, Quebec, Canada: International Computer Music Association.

Legaspi, R., He, Z. & Toyoizumi, T. (2019). Synthetic agency: sense of agency in artificial intelligence. *Current Opinion in Behavioral Sciences*, *29*, 84–90.

Leman, M. (2012). Musical gestures and embodied cognition. In *Journées d'informatique musicale, Proceedings* (pp. 5–7). Université de Mons.

Leman, M., Maes, P.-J., Nijs, L. & Van Dyck, E. (2018). What Is Embodied Music Cognition? In R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, Springer Handbooks (pp. 747–760). Berlin, Heidelberg: Springer.

Leube, D. T., Knoblich, G., Erb, M. & Kircher, T. T. (2003). Observing one's hand become anarchic: An fMRI study of action identification. *Consciousness and Cognition*, *12(4)*, 597–608.

Lewis, G. E. (1996). Improvised Music after 1950: Afrological and Eurological Perspectives. *Black Music Research Journal*, *16(1)*, 91.

Lewis, G. E. (2000). Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, *10(1)*, 33–39.

Li, X., Zhou, P. & Aruin, A. S. (2007). Teager–Kaiser Energy Operation of Surface EMG Improves Muscle Activity Onset Detection. *Annals of Biomedical Engineering*, *35(9)*, 1532–1538.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21(1)*, 1–36.

Linnainmaa, S. (1979). The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master's thesis, University of Helsinki, Finland.

Liontiris, T. P. (2018). Low Frequency Feedback Drones: A non-invasive augmentation of the double bass. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Blacksburg, Virginia, USA: Zenodo.

Locke, J. (1959). *An essay concerning human understanding*. OCLC: 371280.

Lucier, A. (2012). *Music 109: Notes on Experimental Music*. Wesleyan University Press. Google-Books-ID: 0_h90RXiOKwC.

Ludden, G. D. S., Schifferstein, H. N. J. & Hekkert, P. (2008). Surprise as a Design Strategy. *Design Issues*, *24(2)*, 28–38. Publisher: The MIT Press.

Lusted, H. S. & Knapp, R. B. (1988). Biomuse: Musical performance generated by human bioelectric signals. *The Journal of the Acoustical Society of America*, *84(S1)*, S179–S179. Publisher: Acoustical Society of America.

Maes, P. (1993). Modeling Adaptive Autonomous Agents. *Artificial Life*, *1(1_2)*, 135–162.

Maes, P.-J. (2016). Sensorimotor Grounding of Musical Embodiment and the Role of Prediction: A Review. *Frontiers in Psychology*, *7*, 308.

Maes, P.-J., Leman, M., Lesaffre, M., Demey, M. & Moelants, D. (2010). From expressive gesture to sound. *Journal on Multimodal User Interfaces*, *3(1)*, 67–78.

Magnusson, T. (2019). *Sonic writing: technologies of material, symbolic, and signal inscriptions*. OCLC: 1023599945.

Malafouris, L. (2008). At the Potter's Wheel: An Argument for Material Agency. In C. Knappett & L. Malafouris (Eds.), *Material Agency: Towards a Non-Anthropocentric Approach* (pp. 19–36). Boston, MA: Springer US.

Mann, Y. (2016). AI Duet.

Marcus, G. (2018). Deep Learning: A Critical Appraisal.

Marsh, K. L., Richardson, M. J. & Schmidt, R. C. (2009). Social Connection Through Joint Action and Interpersonal Coordination. *Topics in Cognitive Science*, *1(2)*, 320–339. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2009.01022.x.

Martin, C. & Duhaime, D. (2019). cpmpercussion/keras-mdn-layer v0.3.0. https://zenodo.org/record/3526753.

Martin, C. P. (2019). IMPS: Interactive Musical Prediction System: Demo Video. https://zenodo.org/record/2597494.

Martin, C. P., Ellefsen, K. O. & Torresen, J. (2018a). Deep Predictive Models in Interactive Music. *arXiv:1801.10492 [cs, eess]*. arXiv: 1801.10492.

Martin, C. P., Jensenius, A. R. & Torresen, J. (2018b). Composing an Ensemble Standstill Work for Myo and Bela. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 196–197). Blacksburg, Virginia, USA: Zenodo.

Martin, C. P. & Torresen, J. (2018). RoboJam: A Musical Mixture Density Network for Collaborative Touchscreen Interaction. In Liapis, A., Romero Cardalda, J. J. & Ekárt, A. (Eds.), *Computational Intelligence in Music, Sound, Art and Design*, Lecture Notes in Computer Science (pp. 161–176). Springer International Publishing.

Martin, C. P. & Torresen, J. (2019). An Interactive Musical Prediction System with Mixture Density Recurrent Neural Networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 260–265). Porto Alegre, Brazil.

Martins, J. M. & Miranda, E. R. (2007). Emergent Rhythmic Phrases in an A-Life Environment.

Maturana, H. R. (1980). *Autopoiesis and cognition: the realization of the living*, Volume 42 of *Boston studies in the philosophy of science*. Dordrecht: Reidel.

Maturana, H. R. & Varela, F. J. (1980). Cognitive Function in General. In H. R. Maturana & F. J. Varela (Eds.), *Autopoiesis and Cognition: The Realization of the Living*, Boston Studies in the Philosophy and History of Science (pp. 8–14). Dordrecht: Springer Netherlands.

McAlpine, K., Miranda, E. & Hoggar, S. (1999). Making Music with Algorithms: A Case-Study System. *Computer Music Journal*, *23(2)*, 19–30. Publisher: The MIT Press.

McDonald, J. J., Teder-Sälejärvi, W. A. & Ward, L. M. (2001). Multisensory Integration and Crossmodal Attention Effects in the Human Brain. *Science*, *292(5523)*, 1791–1791. Publisher: American Association for the Advancement of Science.

Mcgurk, H. & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264(5588)*, 746–748. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5588 Primary_atype: Research Publisher: Nature Publishing Group.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., Courville, A. & Bengio, Y. (2017). SampleRNN: An Unconditional End-to-End Neural Audio Generation Model. *arXiv:1612.07837 [cs]*. arXiv: 1612.07837.

Melbye, A. P. & Ulfarsson, H. A. (2020). Sculpting the behaviour of the Feedback-Actuated Augmented Bass: Design strategies for subtle manipulations of string feedback using simple adaptive algorithms. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Birmingham, UK: Zenodo.

Mencke, I., Omigie, D., Wald-Fuhrmann, M. & Brattico, E. (2019). Atonal Music: Can Uncertainty Lead to Pleasure? *Frontiers in Neuroscience*, *12*. Publisher: Frontiers.

Mendoza, J. I. & Thompson, M. R. (2017). Gestural Agency in Human–Machine Musical Interaction. In M. Lesaffre, P.-J. Maes & M. Leman (Eds.), *The Routledge Companion to Embodied Music Interaction* (1 Ed.) (pp. 412–419). New York; London: Routledge.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T. & Koelsch, S. (2017). The Distancing-Embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, *40*. Publisher: Cambridge University Press.

Merleau-Ponty, M. (1965). *The structure of behaviour*. London: Methuen.

Merleau-Ponty, M. (2012). *Phenomenology of perception*. London: Routledge.

Michailidis, T., Dooley, J., Granieri, N. & Donato, B. D. (2018). Improvising through the senses: a performance approach with the indirect use of technology. *Digital Creativity*, *29(2-3)*, 149–164.

Miller, L. E. (2001). Cage, Cunningham, and Collaborators: The Odyssey of Variations V. *The Musical Quarterly*, *85(3)*, 545–567.

Mingers, J. (1989). An introduction to autopoiesis—Implications and applications. *Systems practice*, *2(2)*, 159–180.

Minsky, M. (1981). Music, Mind, and Meaning. *Computer Music Journal*, *5(3)*, 28–44. Publisher: The MIT Press.

Minsky, M. (1986). *The society of mind*. USA: Simon & Schuster, Inc.

Miranda, E. R. (2000). *Readings in music and artificial intelligence*, Volume vol. 20 of *Contemporary music studies*. Amsterdam: Harwood Academic.

Miranda, E. R. (2002). Sounds of artificial life. In *Proceedings of the 4th conference on Creativity & cognition* (pp. 173–177). New York, NY, USA: Association for Computing Machinery.

Miranda, E. R. (2011). *A-Life for Music: Music and Computer Models of Living Systems*. A-R Editions, Inc. Google-Books-ID: W2_n1R5F2XoC.

Miranda, E. R. (Ed.). (2021). *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer International Publishing.

Mishra, S., Sturm, B. L. & Dixon, S. (2018). Understanding A Deep Machine Listening Model Through Feature Inversion (p.8̃).

Misselhorn, C. (2015). Collective Agency and Cooperation in Natural and Artificial Systems. In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, Philosophical Studies Series (pp. 3–24). Cham: Springer International Publishing.

Moore, J. W. (2016). What Is the Sense of Agency and Why Does it Matter? *Frontiers in Psychology*, *7*, 1272.

Moore, O. K. (1957). Divination - A New Perspective. *American Anthropologist, 59(1)*, 69–74. Publisher: [American Anthropological Association, Wiley].

Moss, D. (Ed.). (1999). *Humanistic and transpersonal psychology: A historical and biographical sourcebook.* Humanistic and transpersonal psychology: A historical and biographical sourcebook. Westport, CT, US: Greenwood Press/Greenwood Publishing Group. Pages: xxi, 457.

Motl, K. R. (2013). *Multiphonics on the double bass: An investigation on the development and use of multiphonics on the double bass in contemporary music.* PhD thesis, University of California San Diego, San Diego, CA.

Mudd, T., Holland, S. & Mulholland, P. (2019). Nonlinear dynamical processes in musical interactions: Investigating the role of nonlinear dynamics in supporting surprise and exploration in interactions with digital musical instruments. *International Journal of Human-Computer Studies*, *128*, 27–40.

Muller, J. (2019). Stockhausen's "Mikrophonie I". https://jeremymuller.com/mikrophonie/.

Nake, F. (2012). Construction and Intuition: Creativity in Early Computer Art. In J. McCormack & M. d'Inverno (Eds.), *Computers and Creativity* (pp. 61–94). Berlin, Heidelberg: Springer.

Nees, G. & Bense, M. (1965). Projekte generativer Ästhetik. (The projects of generative aesthetics) | Database of Digital Art. *computer-grafikrot.*

Newell, A., Shaw, J. C. & Simon, H. A. (1959). A Variety of Intelligent Learning In A General Problem Solver (p.8̃).

Niewiadomski, R., Mancini, M., Piana, S., Alborno, P., Volpe, G. & Camurri, A. (2017). Low-intrusive recognition of expressive movement qualities. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI '17 (pp. 230–237). New York, NY, USA: Association for Computing Machinery.

Nijs, L., Lesaffre, M. & Leman, M. (2013). The musical instrument as a natural extension of the musician. In *Music and its instruments* (pp. 467–484). Editions Delatour France. ISSN: 2105-908X.

Nort, D. V. (2018). Conducting the in-between: improvisation and intersubjective engagement in soundpainted electro-acoustic ensemble performance. *Digital Creativity, 29(1)*, 68–81.

Nort, D. V., Oliveros, P. & Braasch, J. (2013). Electro/Acoustic Improvisation and Deeply Listening Machines. *Journal of New Music Research, 42(4)*, 303–324.

Noy, L., Dekel, E. & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences*, *108(52)*, 20947–20952. Publisher: National Academy of Sciences Section: Social Sciences.

Noy, L., Levit-Binun, N. & Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Frontiers in Human Neuroscience*, *9*, 187.

Nyman, M. (1999). *Experimental Music: Cage and Beyond*. Cambridge University Press. Google-Books-ID: QEBzEhzAkYwC.

Nymoen, K., Chandra, A. & Torresen, J. (2016). Self-awareness in Active Music Systems. In P. R. Lewis, M. Platzner, B. Rinner, J. Tørresen & X. Yao (Eds.), *Self-aware Computing Systems: An Engineering Approach*, Natural Computing Series (pp. 279–296). Cham: Springer International Publishing.

Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T. & Barabási, A.-L. (2000). The sound of many hands clapping. *Nature*, *403(6772)*, 849–850. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6772 Primary_atype: Research Publisher: Nature Publishing Group.

Oliveros, P. (1984). *Software for people: collected writings 1963-80*. Barrytown, NY: Smith Publications; Printed Editions.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*. arXiv: 1609.03499.

Ortiz, M., Coghlan, N., Jaimovich, J. & Knapp, R. B. (2011). Biosignal-driven Art: Beyond biofeedback. Accepted: 2017-11-29T13:51:16Z Publisher: CMMAS.

Ouzounian, G. (2011). The uncertainty of Experience: On George Brecht's Event Scores. *Journal of Visual Culture*, *10(2)*, 198–211. Publisher: SAGE Publications.

Pachet, F. (2003). The Continuator: Musical Interaction With Style. *Journal of New Music Research*, *32(3)*, 333–341. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1076/jnmr.32.3.333.16861.

Paine, G. (2009). Towards Unified Design Guidelines for New Interfaces for Musical Expression. *Organised Sound*, *14(2)*, 142–155. Publisher: Cambridge University Press.

Paul, E. (2009). Subsonics - Episode 4. https://vimeo.com/3799720.

Peper, E. & Shaffer, F. (2018). Biofeedback History: An Alternative View. *Biofeedback (Online)*, *46(4)*, 80–84. Num Pages: 80-84 Place: Lawrence, United Kingdom Publisher: Allen Press Inc.

Perez-Carrillo, A., Arcos, J.-L. & Wanderley, M. (2016). Estimation of Guitar Fingering and Plucking Controls Based on Multimodal Analysis of Motion, Audio and Musical Score. In Kronland-Martinet, R., Aramaki, M. & Ystad, S. (Eds.), *Music, Mind, and Embodiment*, Lecture Notes in Computer Science (pp. 71–87). Springer International Publishing.

Pfeifer, R. & Bongard, J. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA, USA: A Bradford Book.

Phinyomark, A., Campbell, E. & Scheme, E. (2020). Surface Electromyography (EMG) Signal Processing, Classification, and Practical Considerations. In G. Naik (Ed.), *Biomedical Signal Processing: Advances in Theory, Algorithms and Applications*, Series in BioEngineering (pp. 3–29). Singapore: Springer.

Pierce, J. (1965). Portrait of the Machine as a Young Artist. *Playboy*, *12(6)*, 124–125, 150, 182, 184.

Pinar Saygin, A., Cicekli, I. & Akman, V. (2000). Turing Test: 50 Years Later. *Minds and Machines*, *10(4)*, 463–518.

Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H. & Atzori, M. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. *PLOS ONE*, *12(10)*, e0186132.

Podevijn, G., O'Grady, R., Mathews, N., Gilles, A., Fantini-Hauwel, C. & Dorigo, M. (2016). Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human–swarm interaction. *Swarm Intelligence*, *10(3)*, 193–210.

Poincaré, H. (1914). *Science and Method*. Dover Publications.

Popova, Y. B. & Rączaszek-Leonardi, J. (2020). Enactivism and Ecological Psychology: The Role of Bodily Experience in Agency. *Frontiers in Psychology*, *11*. Publisher: Frontiers.

Prpa, M., Tatar, K., Françoise, J., Riecke, B., Schiphorst, T. & Pasquier, P. (2018). Attending to Breath: Exploring How the Cues in a Virtual Environment Guide the Attention to Breath and Shape the Quality of Experience to Support Mindfulness. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18 (pp. 71–84). New York, NY, USA: ACM. event-place: Hong Kong, China.

Puckette, M. (1985). A real-time music performance system. Technical Report, The MIT Department of Electrical Engineering, Cambridge, MA, USA.

Puckette, M. (1996). Pure Data: another integrated computer music environment. In *in Proceedings, International Computer Music Conference* (pp. 37–41).

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. & Sainath, T. (2019). Deep Learning for Audio Signal Processing. *IEEE Journal of Selected Topics in Signal Processing*, *13(2)*, 206–219.

Puterman, M. L. (1994). Finite-Horizon Markov Decision Processes. In *Markov Decision Processes* (pp. 74–118). John Wiley & Sons, Ltd. Section: 4 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316887.ch4.

Pérez, M. A. O. (2010). *Towards an Idiomatic Compositional Language for Biosignal Interfaces*. doctoral, Queen's University Belfast, New York, New York.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77(2)*, 257–286. Conference Name: Proceedings of the IEEE.

Ramirez, A., Walther, J. B., Burgoon, J. K. & Sunnafrank, M. (2002). Information-Seeking Strategies, Uncertainty, and Computer-Mediated Communication. *Human Communication Research*, *28(2)*, 213–228. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.2002.tb00804.x.

Ranganath, C. & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, *4(3)*, 193–202. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 3 Primary_atype: Reviews Publisher: Nature Publishing Group.

Ranwala, D. (2020). The Evolution of Music and AI Technology.

Rao, A. S. & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. In *Proceedings of the International Conference on Multiagent Systems* (p.8̃). https://www.aaai.org/Papers/ICMAS/1995/ICMAS95-042.pdf.

Reichardt, J. (1968). Cybernetic Serendipity—Getting Rid of Preconceptions. *Studio International*, *176(905)*, 176–77.

Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21(5)*, 188–194.

Roads, C. (1980). Artificial Intelligence and Music. *Computer Music Journal*, *4(2)*, 13–25. Publisher: The MIT Press.

Roads, C. (1985). Research in music and artificial intelligence. *ACM Computing Surveys*, *17(2)*, 163–190.

Roca, M. A. (1994). Epizoo. http://marceliantunez.com/work/epizoo/.

Rogalsky, M. (2010). 'Nature' as an Organising Principle: Approaches to chance and the natural in the work of John Cage, David Tudor and Alvin Lucier. *Organised Sound*, *15(2)*, 133–136. Publisher: Cambridge University Press.

Rosenblatt, F. (1957). The perceptron - A perceiving and recognizing automaton. Technical Report 85-460-1, Cornell Aeronautical Laboratory, Inc., Buffalo, NY.

Rosenboom, D. (1972). Method for Producing Sounds or Light Flashes with Alpha Brain Waves for Artistic Purposes. *Leonardo*, *5(2)*, 141–145. Publisher: The MIT Press.

Rowe, R. (1992). Machine Listening and Composing with Cypher. *Computer Music Journal*, *16(1)*, 43–63. Publisher: The MIT Press.

Rumelhart, D. E. & McClelland, J. L. (1987). Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (pp. 318–362). MIT Press. Conference Name: Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39(6)*, 1161–1178. Place: US Publisher: American Psychological Association.

Russell, S. J. S. J. (2010). *Artificial intelligence : a modern approach*. Third edition. Upper Saddle River, N.J. : Prentice Hall.

Russolo, L., Filliou, R., Pratella, F. B. & Something Else Press (1913). *The art of noise: futurist manifesto*. New York: Something Else Press. OCLC: 676156.

Santello, M., Flanders, M. & Soechting, J. F. (2002). Patterns of Hand Motion during Grasping and the Influence of Sensory Guidance. *The Journal of Neuroscience*, *22(4)*, 1426–1435.

Sarasua, A., Caramiaux, B. & Tanaka, A. (2016). Machine Learning of Personal Gesture Variation in Music Conducting. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 3428–3432). Santa Clara, California, USA: ACM Press.

Sarasúa, , Caramiaux, B., Tanaka, A. & Ortiz, M. (2017). Datasets for the Analysis of Expressive Musical Gestures. In *Proceedings of the 4th International Conference on Movement Computing - MOCO '17* (pp. 1–4). London, United Kingdom: ACM Press.

Sato, A. & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, *94(3)*, 241–255.

Sawyer, R. K. & DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts*, *3(2)*, 81–92.

Schacher, J. C., Miyama, C. & Bisig, D. (2015). Gestural Electronic Music using Machine Learning as Generative Device. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 347–350). Baton Rouge, Louisiana, USA: Zenodo.

Schaeffer, P. (1966). *Traite des objets musicaux: essai interdisciplines.* Paris: Éditions du Seuil. OCLC: 301664906.

Schank, R. C. (1984). *The cognitive computer: on language, learning, and artificial intelligence.* Reading, Mass: Addison-Wesley.

Schiavio, A. (2015). Action, Enaction, Inter(en)action. *Empirical Musicology Review*, *9(3-4)*, 254–262. Number: 3-4.

Schiavio, A. & Høffding, S. (2015). Playing together without communicating? A pre-reflective and enactive account of joint musical performance. *Musicae Scientiae*, *19(4)*, 366–388. Publisher: SAGE Publications Ltd.

Schiavio, A. & Jaegher, H. D. (2017). Participatory Sense-Making in Joint Musical Practice. In *The Routledge Companion to Embodied Music Interaction.* Routledge. Num Pages: 9.

Schillinger, J. (1948). *The Mathematical Basis of the Arts.* New York: Philosophical Library.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 Ed.). Metaphysics Research Lab, Stanford University.

Schmidhuber, J. (2009). Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes. *arXiv:0812.4360 [cs].* arXiv: 0812.4360.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117.

Schneider, A. (2018). Perception of Timbre and Sound Color. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology*, Springer Handbooks (pp. 687–725). Berlin, Heidelberg: Springer.

Schnell, N. & Battier, M. (2002). Introducing Composed Instruments, Technical and Musicological Implications. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 156–160). Dublin, Ireland: Zenodo.

Schnell, N., Röbel, A., Schwarz, D., Peeters, G. & Borghesi, R. (2019). MuBu & Friends - Assembling Tools for Content Based Real-Time Interactive Audio Processing in Max/MSP (p.4̃). Montreal, Quebec, Canada.

Schuster, M. (1999). *On Supervised Learning From Sequential Data With Applications For Speech Recognition.* Ph.D. Thesis, Nara Institute of Science and Technology, Nara, Ikoma, Japan.

Schwarz, K. R. (1980). Steve Reich: Music as a Gradual Process: Part I. *Perspectives of New Music*, *19(1/2)*, 373–392. Perspectives of New Music.

Schwarz, K. R. (1981). Steve Reich: Music as a Gradual Process Part II. *Perspectives of New Music*, *20(1/2)*, 225–286. Perspectives of New Music.

Scurto, H., Van Kerrebroeck, B., Caramiaux, B. & Bevilacqua, F. (2019). Designing Deep Reinforcement Learning for Human Parameter Exploration. *arXiv:1907.00824 [cs, eess]*. arXiv: 1907.00824.

Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10(2)*, 70–76.

Seel, N. (1989). *Agent theories and architectures*. PhD thesis, University of Surrey, London, UK.

Serra, M.-H. (1993). Stochastic Composition and Stochastic Timbre: GENDY3 by Iannis Xenakis. *Perspectives of New Music*, *31(1)*, 236–257. Perspectives of New Music.

Serra, X. (2005). Towards a roadmap for the research in music technology. In *Proceedings of the International Computer Music Conference*.

Shanken, E., Clarke, B. & Henderson, L. (2012). Cybernetics and Art : Cultural Convergence in the 1960s.

Shannon, C. E. (1950). XXII. Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *41(314)*, 256–275.

Shany, O., Singer, N., Gold, B. P., Jacoby, N., Tarrasch, R., Hendler, T. & Granot, R. (2019). Surprise-related activation in the nucleus accumbens interacts with music-induced pleasantness. *Social Cognitive and Affective Neuroscience*, *14(4)*, 459–470.

Shapiro, L. (2010). *Embodied Cognition* (2 Ed.). New problems of philosophy. Milton: Routledge.

Shoemaker, S. (1984). *Identity, cause, and mind: philosophical essays*. Cambridge: University Press.

Shultz, T. R. (1991). From agency to intention: A rule-based, computational approach. In *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 79–95). Cambridge, MA, US: Basil Blackwell.

Simon, H. A. (1996). *Models of my life*. London: MIT Press.

Skogstad, S. A. v. D. (2014). Methods and Technologies for Using Body Motion for Real-Time Musical Interaction.

Slotnick, D. L. (1971). The Fastest Computer. *Scientific American*, *224(2)*, 76–87. Publisher: Scientific American, a division of Nature America, Inc.

Smalley, D. (1997). Spectromorphology: explaining sound-shapes. *Organised Sound*, *2(2)*, 107–126.

Smith (1980). The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computers*, *C-29(12)*, 1104–1113. Conference Name: IEEE Transactions on Computers.

Smith, B. D. & Garnett, G. E. (2012). Unsupervised Play: Machine Learning Toolkit for Max. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. x–x). Ann Arbor, Michigan: Zenodo.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11(1)*, 1–23. Publisher: Cambridge University Press.

Snyder, J. & Ryan, D. (2014). The Birl: An Electronic Wind Instrument Based on an Artificial Neural Network Parameter Mapping Structure. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 585–588). London, United Kingdom: Zenodo.

Spence, A. (2016). Intonarumori. https://alexandraspence.net/Intonarumori-the-VAG.

St-Amant, Y., Rancourt, D. & Clancy, E. (1996). Effect of smoothing window length on RMS EMG amplitude estimates. In *Proceedings of the IEEE 22nd Annual Northeast Bioengineering Conference* (pp. 93–94). ISSN: null.

St-Onge, D., Côté-Allard, U., Glette, K., Gosselin, B. & Beltrame, G. (2019). Engaging with Robotic Swarms: Commands from Expressive Motion. *ACM Transactions on Human-Robot Interaction*, *8(2)*, 11:1–11:26.

Stelarc (1980). Third Hand.

Straebel, V. & Thoben, W. (2014). Alvin Lucier's Music for Solo Performer: Experimental music beyond sonification. *Organised Sound*, *19(1)*, 17–29. Cambridge University Press.

Strasser, A. (2015). Can Artificial Systems Be Part of a Collective Action? In C. Misselhorn (Ed.), *Collective Agency and Cooperation in Natural and Artificial Systems: Explanation, Implementation and Simulation*, Philosophical Studies Series (pp. 205–218). Cham: Springer International Publishing.

Straussfogel, D. & von Schilling, C. (2009). Systems Theory. In R. Kitchin & N. Thrift (Eds.), *International Encyclopedia of Human Geography* (pp. 151–158). Oxford: Elsevier.

Strickland, S. (2018). Ringing the Changes: Mapping the Algorithmic Art of Change-Ringing on Church Bells. *Ecotone*, *13(2)*, 138–144. Publisher: University of North Carolina, Wilmington.

Sturm, B., Santos, J. F. & Korshunova, I. (2015). Folk music style modelling by recurrent neural networks with long short term memory units. In *16th International Society for Music Information Retrieval Conference, late-breaking demo session*.

Synofzik, M., Thier, P. & Lindner, A. (2006). Internalizing Agency of Self-Action: Perception of One's Own Hand Movements Depends on an Adaptable Prediction About the Sensory Action Outcome. *Journal of Neurophysiology*, *96(3)*, 1592–1601. Publisher: American Physiological Society.

Tahiroglu, K., Kastemaa, M. & Koli, O. (2021). AI-terity 2.0: An Autonomous NIME Featuring GANSpaceSynth Deep Learning Model. In *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Takayama, L. (2012). Perspectives on Agency Interacting with and through Personal Robots. In M. Zacarias & J. V. de Oliveira (Eds.), *Human-Computer Interaction: The Agency Perspective*, Studies in Computational Intelligence (pp. 195–214). Berlin, Heidelberg: Springer.

Tanaka, A. (1993). Musical Technical Issues in Using Interactive Instrument Technology with Application to the BioMuse. In *Proceedings of the International Computer Music Conference*.

Tanaka, A. (2000). Musical performance practice on sensor-based instruments. *Trends in Gestural Control of Music*, *13(389-405)*, 284.

Tanaka, A. (2011). Sensor-Based Musical Instruments and Interactive Music. In *The Oxford Handbook of Computer Music* (Roger T. Dean Ed.). Oxford University Press.

Tanaka, A. (2015). Intention, Effort, and Restraint: The EMG in Musical Performance. *Leonardo*, *48(3)*, 298–299.

Tanaka, A. & Donnarumma, M. (2018). The Body as Musical Instrument. *The Oxford Handbook of Music and the Body*.

Tani, J. (1998). An Interpretation of the "Self" From the Dynamical Systems Perspective: A Constructivist Approach. *Journal of Consciousness Studies*, *5*, 516–542.

Tatar, K., Bisig, D. & Pasquier, P. (2020). Introducing Latent Timbre Synthesis. *arXiv:2006.00408 [cs, eess]*. arXiv: 2006.00408.

Tatar, K. & Pasquier, P. (2017). MASOM: A Musical Agent Architecture based on Self-Organizing Maps, Affective Computing, and Variable Markov Models (p.8̃). Atlanta, United States: MuMe.

Tatar, K. & Pasquier, P. (2019). Musical agents: A typology and state of the art towards Musical Metacreation. *Journal of New Music Research*, *48(1)*, 56–105.

Thelen, E. (1996). Time-scale dynamics and the development of an embodied cognition. In *Mind as motion: explorations in the dynamics of cognition* (pp. 69–100). USA: Massachusetts Institute of Technology.

Thelen, E., Schöner, G., Scheier, C. & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24(1)*, 1–34. Publisher: Cambridge University Press.

Think, D. (2019). Noise Mixer. https://dont-think.bandcamp.com/album/noise-mixer.

Thom, B. (2000). BoB: an interactive improvisational music companion. In *Proceedings of the fourth international conference on Autonomous agents - AGENTS '00* (pp. 309–316). Barcelona, Spain: ACM Press.

Thom, B. (2001). Interactive Improvisational Music Companionship: A User-Modeling Approach (p.5̃).

Tienson, J. L. (1987). Introduction to Connectionism. *Southern Journal of Philosophy (Suppl.)*, *1*, 1–16.

Todd, P. M. (1989). A Connectionist Approach to Algorithmic Composition. *Computer Music Journal*, *13(4)*, 27–43. The MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28(5)*, 675–691. Cambridge University Press.

Turing, A. (1950). Computing, Machinery and Intelligence. *Mind*, *LIX(236)*, 433–460.

Tuuri, K., Parviainen, J. & Pirhonen, A. (2017). Who Controls Who? Embodied Control Within Human–Technology Choreographies†. *Interacting with Computers*, *29(4)*, 494–511.

Uliam, H., de Azevedo, F. M., Ota Takahashi, L. S., Moraes, E., Negrao Filho, R. d. F. & Alves, N. (2012). The Relationship Between Electromyography and Muscle Force. In M. Schwartz (Ed.), *EMG Methods for Evaluating Muscle and Nerve Function*. InTech.

Vacuo, F. (2020). Humane Methods.

Van Nort, D., Wanderley, M. M. & Depalle, P. (2014). Mapping Control Structures for Sound Synthesis: Functional and Topological Perspectives. *Computer Music Journal*, *38(3)*, 6–22.

Varela, F. G., Maturana, H. R. & Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. *Currents in Modern Biology*, *5(4)*, 187–196.

Varela, F. J., Thompson, E. & Rosch, E. (1991). *The embodied mind: cognitive science and human experience*. Cambridge, Mass: MIT Press.

Vaucanson, J. D. (2018). *An Account of the Mechanism of an Automaton, or Image Playing on the German-Flute: As It Was Presented in a Memoire, to the Gentlemen of the ... with a Description of an Artificial Duck,*. Place of publication not identified: Gale Ecco, Print Editions.

Viglione, S. S. (1970). 4 Applications of Pattern Recognition Technology. In J. M. Mendel & K. S. Fu (Eds.), *Mathematics in Science and Engineering*, Volume 66 of *Adaptive, Learning and Pattern Recognition Systems* (pp. 115–162). Elsevier.

Visi, F. G. & Tanaka, A. (2020). Towards Assisted Interactive Machine Learning: Exploring Gesture-Sound Mappings Using Reinforcement Learning (p.1̃0). Trondheim, Norway.

Visi, F. G. & Tanaka, A. (2021). Interactive Machine Learning of Musical Gesture. In E. R. Miranda (Ed.), *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (pp. 771–798). Cham: Springer International Publishing.

Vogt, E. Z. (1959). *Water Witching U.S.A.* Chicago: The University of Chicago Press.

Wallace, B., Martin, C. P., Torresen, J. & Nymoen, K. (2020). Towards Movement Generation with Audio Features. *arXiv:2011.13453*.

Wan, A. D. M. & Braspenning, P. J. (1996). Agent Theory: Autonomy and Self-Control. In *The 7th International Conference on Artificial Intelligence: Methodology, Systems and Applications (AIMSA 96)* (pp. 268–277).

Ward, N., Ortiz, M., Bernardo, F. & Tanaka, A. (2016). Designing and measuring gesture using laban movement analysis and electromyogram. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16* (pp. 995–1000). Heidelberg, Germany: ACM Press.

Watakabe, M., Mita, K., Akataki, K. & Itoh, Y. (2001). Mechanical behaviour of condenser microphone in mechanomyography. *Medical & Biological Engineering & Computing*, *39(2)*, 195–201.

Wegner, D. M. (2002). *The illusion of conscious will*. London: Bradford Books.

Wegner, D. M. (2003). The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences*, *7(2)*, 65–69.

Wegner, D. M. & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54(7)*, 480. Publisher: US: American Psychological Association.

Weinberg, G., Bretan, M., Hoffman, G. & Driscoll, S. (2020a). *Robotic Musicianship: Embodied Artificial Creativity and Mechatronic Musical Expression*. Automation, Collaboration, & E-Services. Springer International Publishing.

Weinberg, G., Bretan, M., Hoffman, G. & Driscoll, S. (2020b). "Wear it"—Wearable Robotic Musicians. In G. Weinberg, M. Bretan, G. Hoffman & S. Driscoll (Eds.), *Robotic Musicianship: Embodied Artificial Creativity and Mechatronic Musical Expression*, Automation, Collaboration, & E-Services (pp. 213–254). Cham: Springer International Publishing.

Welch, D. & Fremaux, G. (2017). Why Do People Like Loud Sound? A Qualitative Study. *International Journal of Environmental Research and Public Health*, *14(8)*, 908. Multidisciplinary Digital Publishing Institute.

van der Wel, R. P., Sebanz, N. & Knoblich, G. (2012a). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*, *21(3)*, 1267–1279.

van der Wel, R. P. R. D., Sebanz, N. & Knoblich, G. (2012b). Action Perception from a Common Coding Perspective. In K. Johnson & M. Shiffrar (Eds.), *People Watching* (pp. 101–118). Oxford University Press.

Whitehead, A. N. (1910). *Principia mathematica*. Nineteenth Century Collections Online (NCCO): Science, Technology, and Medicine: 1780-1925. Cambridge: University Press.

Whitelaw, M. (2004). *Metacreation: Art and Artificial Life*. Cambridge, MA, USA: MIT Press.

Wiener, N. (1948). *Cybernetics; or control and communication in the animal and the machine*. Cybernetics; or control and communication in the animal and the machine. Oxford, England: John Wiley. Pages: 194.

Wiering, M. & van Otterlo, M. (Eds.). (2012). *Reinforcement Learning*, Volume 12 of *Adaptation, Learning, and Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Wishart, T. (1996). *On sonic art* (A new and rev. ed. edited by Simon Emmerson. Ed.), Volume vol. 12 of *Contemporary music studies*. Amsterdam: Harwood Academic Publ.

Wittgenstein, L. (1969). *Preliminary studies for the 'Philosophical investigations',: Generally known as the Blue and Brown books* (2nd edition Ed.). Oxford: Blackwell.

Wolpert, D. M., Ghahramani, Z. & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269(5232)*, 1880–1882. Publisher: American Association for the Advancement of Science.

Woodworth, R. S. (1939). Individual and Group Behavior. *American Journal of Sociology*, *44(6)*, 823–828. Publisher: University of Chicago Press.

Wooldridge, M. & Jennings, N. R. (1995). Agent theories, architectures, and languages: A survey. In Wooldridge, M. J. & Jennings, N. R. (Eds.), *Intelligent Agents*, Lecture Notes in Computer Science (pp. 1–39). Berlin, Heidelberg: Springer.

Zandt-Escobar, A. V., Caramiaux, B. & Tanaka, A. (2014). PiaF: A Tool for Augmented Piano Performance Using Gesture Variation Following. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. London, United Kingdom: Zenodo.

Zhang, X.-D. (2020). Machine Learning. In X.-D. Zhang (Ed.), *A Matrix Algebra Approach to Artificial Intelligence* (pp. 223–440). Singapore: Springer.

Zicarelli, D. (1987). M and Jam Factory. *Computer Music Journal*, *11(4)*, 13–29. The MIT Press.

Zicarelli, D. D. (1998). An Extensible Real-Time Signal Processing Environment for MAX. In *Proceedings of the International Computer Music Conference* (p.4̃). Beijing, China.

Papers

Paper I

Vrengt: A Shared Body–Machine Instrument for Music–Dance Performance

Çağrı Erdem, Katja Henriksen Schia, Alexander Refsum Jensenius

Published in *Proceedings of the 14th International Conference on New Interfaces for Musical Expression*, pp. 186–191.
June 2019

Vrengt: A Shared Body–Machine Instrument for Music–Dance Performance

Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
Department of Musicology
University of Oslo
cagri.erdem@imv.uio.no

Katja Henriksen Schia
PRAXIS
Norwegian Contemporary Dance Company
katjaschia@gmail.com

Alexander Refsum Jensenius
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
Department of Musicology
University of Oslo
a.r.jensenius@imv.uio.no

ABSTRACT

This paper describes the process of developing a shared instrument for music–dance performance, with a particular focus on exploring the boundaries between standstill vs motion, and silence vs sound. The piece *Vrengt* grew from the idea of enabling a true partnership between a musician and a dancer, developing an instrument that would allow for active co-performance. Using a participatory design approach, we worked with sonification as a tool for systematically exploring the dancer's bodily expressions. The exploration used a "spatiotemporal matrix," with a particular focus on sonic microinteraction. In the final performance, two Myo armbands were used for capturing muscle activity of the arm and leg of the dancer, together with a wireless headset microphone capturing the sound of breathing. In the paper we reflect on multi-user instrument paradigms, discuss our approach to creating a shared instrument using sonification as a tool for the sound design, and reflect on the performers' subjective evaluation of the instrument.

Author Keywords

Music, dance, EMG, breathing, sonification, sound synthesis, multi-user instruments, comprovisation

CCS Concepts

•**Applied computing** → **Sound and music computing;** Performing arts; •**Human-centered computing** → *User centered design;*

1. INTRODUCTION

In today's experimental performance scene, many musicians are exploring performance practices that approach dance, and many dancers are working with interactive music systems. A challenge in such exploration, however, is fundamentally different intentions ranging from particular embodied practices [36]. For a musician, the sound is the primary focus of attention, and the movements needed to produce the sound (the sound-producing and sound-modifying actions) are the result of that aim. For a dancer, on the other hand, the movements are the primary focus of attention, and any sonic output is secondary. It is therefore not surprising that the dancer in an interactive context does not intuitively render her movements into instrumental actions for active sound-making, but rather maintains her regular dance-actions influencing the sound generation in an abstract way. Similarly, the musician either takes the role of the composer without active involvement, or, as the performer enacting her own instrument.

Figure 1: The dancer, blindfolded, in the first live performance of *Vrengt*. (Photo: Sophie C. Barth)

In this paper, we continue our exploration of working between dance and music, this time focusing on co-performance on a "shared" instrument. As opposed to creating a system for interactive dance, we wanted to develop what is experienced as one, coherent instrument that enables a true partnership for the musician and dancer. The challenge, then, is to what extent the dancer is able to adopt musical intentions on top of her movement practice, and whether the composer–performer can waive the control of performing while still "playing together"?

2. BACKGROUND

2.1 Between the conscious and the unconscious

Experiencing the body as part of your subjective presence rather than a mere series of shapes on the stage, is described by dancers as "being in your body" [34]. This is often the result of skill acquisition, which Dreyfus has argued is a con-

NIME'19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

tinuum of less and less processing information at a cognitive level [7]. In other words, we operate more intuitively and less consciously as we gain expertise. In music, such skill acquisition is often based on proprioceptive relationships between a musician and instrument [31]. In fact, most human movement is found in the span between conscious and unconscious. That is, we unconsciously execute a number of physiological and biological processes for a single, deliberate task [3]. This is something that has been explored in the context of music–dance performances under the labeling sonic microinteraction [18, 16].

2.2 Multi-user Instruments

Multi-user instruments have become more popular in recent years, but this is still a fairly unexplored territory. Historically, there are several examples of shared musical instrument practice, in particular in the form of four-handed piano works from the 18th and 19th centuries [13]. At that time, the shared performance allowed for forced intimacy in a social space, serving also as away of bridging the gaps of skills and social grades [5]. In the 20th century, experimental composers, such as John Cage and Karlheinz Stockhausen, explored the musical possibilities gained by exploiting the complex relationships between multiple users [19]. But it was first with digital technologies that the idea of designing instruments specifically to work together on and around the same musical content took off [20]. Some notable examples from the NIME community include the Tooka [10] and Reactable [21], and a number of more recent web-based instruments may also be classified as multi-user.

2.3 Interactive Dance

The second author is proficient in release-based training, which is a contemporary dance technique that focuses on performing tasks with least amount of muscle exertion by using the gravity [25]. A challenge in an interactive dance context is to design an interface that allows the dancer control of the sound, but without sacrificing the existing performance technique [38]. It is particularly important to allow for *flow* procedures, in which there is an immediate and causal feedback, yet at the same time a "sense of discovery" [4]. For that reason we have been interested in using sonification as a tool, since it is often thought of as a more "objective" approach to rendering sound in response to data than more creatively based sound design [15]. There are numerous examples of the use of sonification in dance-related motion analysis [28], dance pedagogy and education [11, 14], supporting the development of interactive dance pieces [23, 18] as well as assisting dancers with disabilities [22, 29]. In our case the sonification is not the end result, but rather a tool used as part of the creative process.

3. CONCEPTUAL DESIGN

The main idea of *Vrengt* was that of creating a body–machine instrument in which the dancer would interact with her body and the musician with a set of physical controllers. As such, it may seem as a quite normal setup for a music–dance performance, except that we did not want the dancer and musician to work in separate "layers," but rather co-control the same sonic and musical parameters. This was conceptually different than they had done before. The development was done using a participatory design approach, combining a series of analyses, conversations, recording sessions, and subjective evaluations during the development of the instrument and final performance. As such, the entire process was very integrated, and both the musician and dancer felt a complete ownership of the final "product."

3.1 Interaction Concept

Our project grew from the concept of human micromotion, the tiniest producible and observable motion. These can be used in sonic microinteraction, which are found in most performances on acoustic instruments, but arguably not so often in digital musical instruments [17]. We start from capturing the "smallest components" of the dancer's bodily exertions in the form of muscle signals and breathing, explore them through sonification, and then gradually build the entire system up from there.

Electromyogram (EMG) is a complex signal that represents the electrical currents generated during neuromuscular activities. It is able to report little or non-visible "inputs" (*intentions*), which may not always result in overt body movements [43]. EMG is therefore highly relevant for exploring involuntary micromotion. The first author has been exploring what a muscle interface can add to the existing interaction paradigms of traditional instrumentalists [9]. "Playing with muscles" can enhance the engagement with the instrument [30], which should be considered at the top of the design hierarchy [32].

3.2 Compositional Structure

The performance of *Vrengt* may be seen as a *comprovisation* [8], in which the "composed" aspect of the instrument and choreography provides a large amount of freedom in collectively exploring sonic interactions throughout the performance. The piece was structured in three parts:

1. Breath: The first part explores the embodied sounds of the dancer. Her face is covered (Figure 1), which physically forces her to leverage the kinesthetic and auditory senses. She explores the creation of acoustic feedback loops based on the proximity to the speakers, and these loops are modulated and dynamically controlled by the musician.

2. Standstill: This section exploits using micromotion in sonic microinteraction. The dancer describes standing still as "registering 'what is happening' inside my body without the need of moving, which also introduces the gravity, meditation and body-awareness." Even though her micromotion is barely visible, the audience gradually starts to hear the direct audification of the dancer's varying neural commands leading to muscle contraction.

3. Musicking: Both performers join the active process of music-making. With the dancer's own words, this is where she is "accessing the musician's skills and vice versa." During the first two sections, the audience becomes accustomed with the improvised movement patterns; the relationship between these movements and the variations in breath patterns; how her tiniest bodily exertions "sound" during standstill; and finally, how these sounds evolve throughout as she gradually switches from *StandStill* to *Musicking*.

4. IMPLEMENTATION

The hardware system of *Vrengt* includes (Figure 2):

- two Myo armbands, one placed on left forearm and one on the right calf muscle of the dancer
- a wireless headset microphone
- a MIDI controller
- two laptop computers running Max/MSP patches

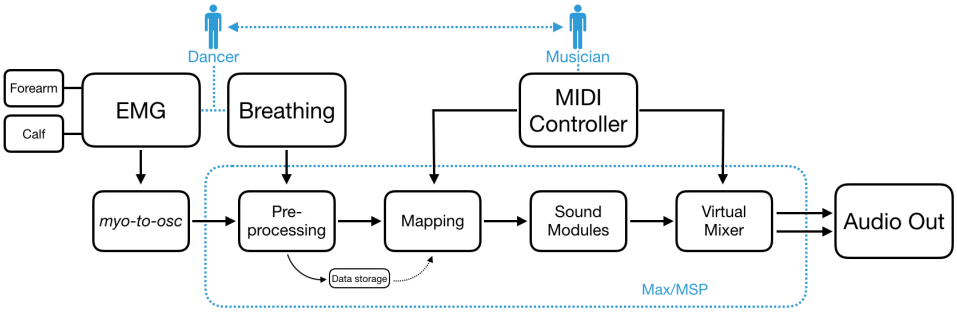


Figure 2: Signal flow diagram for the performance of *Vrengt*.

The armbands are connected to the computer via individual Bluetooth Low Energy (BLE) adapters for overcoming possible bandwidth limitations. The EMG data is acquired with Myo Armband's fixed sample rate at 200Hz and sent via *myo-to-osc* [26] into Max, where the raw EMG signals are pre-processed for full-wave rectification, smoothing and feature extraction (Figure 2).

A lightweight and unobtrusive head-worn condenser lavalier microphone (Sennheiser SL Headmic) is used for capturing the breathing in the form of audio signals that are sent through a wireless transmitter to the laptops.

4.1 Mapping

Inspired by the *spatiotemporal matrix* [17], we started the mapping exploration by recording raw EMG time series of the dancer's muscle activity at different levels: *micro* (during standstill), *meso* (finger extensions, arm flexion) and *macro* (larger actions). Then we tried various feature extractors from [33], among which we decided to use the *mean absolute value* (MAV) upon a preliminary evaluation by mapping the processed data into the sound objects. After having defined the basic structure of the mappings, we subjectively evaluated each distinct action through a process of "cross-modal interpretation." The dancer performed the given patterns mapped into different sound objects, and described her experiences figuratively, in order to determine the meaningful action–sound causalities (Table 1).

Our exploration of perception–action relationships may be seen as unnecessarily time-consuming, but we found this to be necessary to better understand "what is happening" between the body and the sound. This is often "arcane" information embedded in the computational processes. The main user interface for the purpose of shared control is a custom virtual mixer that sums the individual sound modules, allowing the musician to modify the mix levels of the resultant sounds (volume, panning, effects, and so on) along with the data processes (e.g. routing and feature scaling). This is inspired by the seminal work of Alvin Lucier's *Music for Solo Performer* (1965), in which his assistants controlled the sound modules throughout the performance [42].

4.2 Sound Objects

Physics-based synthesis simulates acoustic excitation and resonance features [40, 12] to approximate responsive physical behaviors in digital domain [35], particularly for continuous physical interaction [27]. In our work we have used the *Sound Design Toolkit (SDT)* for physically–informed procedural sound synthesis in Max, specifically the low–level models (e.g. *friction* and *bubble*) and complex textures (e.g. *scraping* and *fluidflows*) [1]. These have been

Table 1: Sonic imagery of mapped relationships

Body Motion	Sound Object	Perceived Sensations
Standing still	*Friction*	"Planting deeply"
Walking	*Friction*	"Squeaking"
Finger flexion	*FluidFlows*	"Squeezing a wet sponge"
Wrist extension	*FluidFlows*	"Casting a fishing line"
Abduction	*Scraping*	"Expanding like a balloon"
Adduction	*Scraping*	"To deflate"
Various	*Waveshaping*	"Shapes without images"

combined with the effects processing objects (e.g. *scrub~* and *pit_shift~*) from the *PeRColate* collection [45].

The *SDT* basic solid interactions are based on a modular "resonator–interactor–resonator" structure [1]. This allows a fairly straightforward thinking in building mapping strategies that refer to physical phenomena between objects in contact. The sound of friction, for instance, is a phenomenon that is most often present in our lives [37], such as the sound of a squeaking door or a knife sliding on a ceramic plate. We can then imagine several "meaningful" ways of associating body movements with everyday sounds.

We used *many-to-many* mappings between the calf muscle signals and the force, pressure, stiffness, dissipation and velocity parameters of the interactor algorithm (*sdt.friction~*), together with the center frequency of a narrow-Q band pass filter, to provide us with a sense of "squeaking" in the motions of the lower limb. Similarly, we used force, grain and velocity parameters of *sdt.scraping~* to evoke the feeling of "filing" when moving the upper limb. However, the perceived sense of the latter model was quite different in the end (see Table 1).

In liquids, sounds are heard only when the air is trapped by water [24]. It is therefore a convenient approach to draw on the acoustical properties of bubbles when designing interactions with liquid sounds. A single, impulsive bubble sound is defined by its radius and rising factor (*ibid*), which is simulated by exponentially decaying sinusoidal oscillators [1]. Then, more complex phenomena can be obtained through statistical approaches as in the *sdt.fluidflow~*, which is a stochastic model. Our strategy was employing the signals of the forearm muscle to modify the speed, density and radii of a stream of bubbles, together with the amount of *scrub~* delay [45] for spatial enhancement. This provided us with sounds that can dynamically morph back and forth, in a continuum between rhythm and tone, echoing the *unified time structuring* of Stockhausen [41].

Additionally, we have explored non-linear (abstract) techniques, such as waveshape distortion, ring modulation (RM) and exponential frequency modulation (FM) for textural purposes. One technique we found intuitive, was to exponentially re-scale the sine wave carrier with multiple sine modulators in a continuous manner through several *many-to-many* mappings that are also exponentially and randomly re-scaled. The result is a quasi-stochastic behavior resembling some of the non-linearities found in using extended techniques on acoustic instruments [43].

For the breath signals, we have implemented *Schroeder Reverberators* [39] together with interconnected multiple delay lines, particularly for sustaining fast attacks. These are simultaneously controlled by the musician, allowing the dancer to interact with the physical space via intentional acoustic feedback loops.

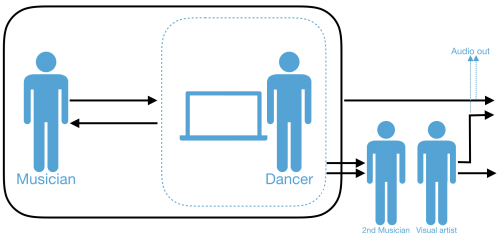


Figure 3: The setup for the final collaborative performance, showing the levels of connection between performers and instruments.

5. DISCUSSION

Vrengt has been performed twice in public so far, once on stage in a large auditorium, and another time in a club environment. In the latter it was performed together with an additional musician and a visual artist. This showed how we can use the instrument in further collaborative situations.[1] In the following, we briefly discuss some of the thoughts we have had during those processes, specifically the subjective evaluations of the dancer and the musician.

5.1 Musician

For a traditionally trained musician and composer to start working with interactive dance, requires stepping outside the comfort zone. Years of experience with working within a familiar instrumental paradigm has to be exchanged with imagining oneself in the athletic and artistic circumstances of a dancer. This was the reason we decided to embark on a fairly long, exploratory journey of the dancer's movement patterns: from involuntary micromotions to deliberate full body movements. The analyses of the sensor data was followed by a number of trials during which different sound objects provided the musician with an experience-based schemata for evaluating the ecological validity of action–sound causalities (see Section 4) and particular sound synthesis models.

The second part of the development involved rehearsals[2] and verbal communication to start shaping the sound design. This phase also involved developing a shared language for describing the experience, using metaphors such as "squeezing a wet sponge" for grasping finger motion, or "planting deeply" for standing still. Such comments are necessary to understand the dancer's feelings, despite the lack of haptic experience when performing in the air. Moreover, such comments are powerful enough to define a path for future work on the relevant topics of sonic interaction design.

Figure 3 describes how the musician sees and experiences the system. The dancer is the main source of gestural input, but the musician makes the decisions of the sound objects, data scaling, and mix levels in realtime. This influences and steers the dancer who, in her own words, "moves through listening." In fact, from the musician's perspective, one can draw an analogy between the dancer and the autonomous musical agents of generative systems. In this sense, the "genericity" of the dancer leans towards the right end of *the continuum of autonomy* in [44], as she learns how to interact with the musician.

The presence of the musician in this project is enacted by means of the dancer's autonomy together with the machine's data processing and sound generation abilities. This echoes the notion of "shared control" in the field of robotic musicianship, which often implies machine intelligence that augments human capabilities [2]. The purpose of such an analogy is not to get into a debate about the human versus the machine, but rather to portray the intimacy between the dancer's body and the machine, and how that is shared by the musician.

5.2 Dancer

From the dancer's perspective, performing with realtime sonification is fundamentally different than dancing to music. In the former case, the sonification steers the movements at both conscious and unconscious levels, and provides a sense of coherence. However, in the latter case, you may experience a "less or unpredictable sense of coherence." Throughout the collaboration, the potential of gesture–sound relationships became more clear, which allowed the dancer to develop "a gestural repertoire and a physical landscape" with a sophisticated control of her movement, and hence sound. This enabled listening as the main source for decision making, while intuitively moving along with "a physical play and exploration." An interesting way of how she portrays her experience with the gained ability of sound-producing is as "a duet" of movement and the sound. As she puts it:

> "The precision between the muscle activation and listening drives the duet forward. It is like the ability to enter a state of not knowing where to, and how to, still with a clear sense of direction. To uncover specificity in the field of movement and sound; making sense collectively to hear the dance and to embody the sound."

One satisfactory aspect of such an instrument from the dancer's perspective, is the shift of focus from the body to the sound. This is described by the dancer as "the sensation of moving through listening," which echoes Paine's *techno-somatic dimension* [32]. In addition to the "feeling" of playing on the instrument, she indicates how her experience with the "sonified muscle tension" resembles her use of tactility when an oral explanation is insufficient. She describes her experience of working with muscle signals as:

> "Learning to relate to a new type of body and a new physical language that can provide an audible response."

[1] Video available at `https://youtu.be/hpECGAkaBp0`
[2] Excerpts of video footage from rehearsals can be seen at `https://bit.ly/2CKl5Ia`

Figure 4: The dancer, the musician and the second musician in a rehearsal.

Her impressions about co-performing on the same instrument is described as "playing together while accessing each other's skills." She uses the Norwegian word "vrengt," which she exemplifies as the act of turning a sweater outwards, pointing to how the artistic intentions and skills are merged together. Furthermore, she emphasizes how each different sound object has a distinct image in her mind (see Table 1) and she "examined the duration, pace and consistency of every movement within them." Reflecting on the use of abstract algorithms for sound synthesis, she comments that they resemble shapes that she can "fill with any image you want." This can be seen as opposed to more straightforward sonic imagery of physics-based models. Moreover, she cannot choose one or the other technique in terms of the level of engagement and embodied control. It is an important user-centered aspect, which should be further investigated.

5.3 A "Shared" Reflection

The usefulness of *Vrengt*'s shareability to the overall aesthetics can be discussed in terms of the unity of two bodies and two machines. This relates to how Marco Donnarumma conceptualizes human-machine embodiment as "a form of hybrid corporeality where experience, psyche, materiality and technics are always in tension against each other" [6]. A natural outcome of this hybrid embodiment is an intimate, bodily knowledge of each other at the boundary between cognitive vs unconscious. This is different than sharing the same stage while not in a joint technological configuration.

We observe the first aesthetic consequence of this unity in the *Breath* part of the piece. What makes the role of the musician different than a tonmeister in controlling the acoustic feedback loops (see Section 3.2) is the multidimensional knowledge of the dancer's breath patterns. At the other end, the musician's interactions become part of how the dancer's bodily exertions happen to be in a sound-producing context. Thus, the overall aesthetics can be viewed as an n-dimensional space of bodily and technical co-dependencies.

Similar forms of co-dependence are observed in the *Standstill* and *Musicking* sections. These forms are based on the ongoing complex bodily interactions at various spatial, physiological and cognitive levels. We can then argue that the particular aesthetic results of *Vrengt* would not have been achieved with other methods, such as working in separate and/or fixed layers.

Perhaps the most significant issue in conceptualizing *Vrengt* as a multi-user instrument, is the performers' uneven bodily contributions. In a more balanced scenario, the musician would use a sensor-based controller, thereby creating more of a hybrid corporeality. In our current setup, the shareability of *Vrengt* is at the musician's "fingertips" only, when compared to the dancer's full-body experience.

6. CONCLUSIONS

In this paper, we have presented the development of a multi-user instrument used in a music–dance performance context. This project has been centered on a common apparatus, in which shareability, sonification, micromotion, and muscle activity have been core elements. We have aimed to design a shareable instrument that blends distinct embodied skills. The final result is a joint musical expression of two performers. This has been achieved by building an entirely situated design methodology, starting from investigating the dancer's breathing and other involuntary micromotion while standing still. This was followed by using sonification as an artistic-scientific tool to explore and enhance the data in question. Furthermore, using various physics-based and abstract sound synthesis techniques allowed for subjectively evaluating their cross-modal associations and levels of embodiment.

In future research, we will continue to build on the model of shared agency developed for *Vrengt*. We are particularly interested in exploring the body as a musical interface. This will be done with a particular focus on the co-creativity of humans and machines, and using intuitive control strategies for physical modeling synthesis and embodied sonic cognition.

7. ACKNOWLEDGMENTS

We would like to thank to Qichao Lan, who collaborated the project as the second musician, and Victor Evaristo Gonzalez Sanchez for his continuous support and comments throughout the development process. This work was partially supported by the Research Council of Norway (project 262762) and NordForsk (project 86892).

8. REFERENCES

[1] S. Baldan, S. Delle Monache, and D. Rocchesso. The sound design toolkit. *SoftwareX*, 6:255–260, 2017.

[2] M. Bretan, D. Gopinath, P. Mullins, and G. Weinberg. A robotic prosthesis for an amputee drummer. *arXiv preprint arXiv:1612.04391*, 2016.

[3] D. Chi, M. Costa, L. Zhao, and N. Badler. The emote model for effort and shape. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 173–182, 2000.

[4] M. Csikszentmihalyi. The flow of creativity. In *Creativity: Flow and The Psychology of Discovery and Invention*, pages 95–112, New York, 1996. Harper Collins.

[5] A. Daub. *Four-handed monsters: four-hand piano playing and nineteenth-century culture*. Oxford University Press, 2014.

[6] M. Donnarumma. Beyond the cyborg: performance, attunement and autonomous computation. *International Journal of Performance Arts and Digital Media*, 13(2):105–119, 2017.

[7] H. L. Dreyfus. Phenomenological description versus rational reconstruction. *Revue internationale de philosophie*, (2):181–196, 2001.

[8] R. Dudas. "comprovisation": The various facets of composed improvisation within interactive

performance systems. *Leonardo Music Journal*, pages 29–31, 2010.

[9] C. Erdem, A. Camci, and A. Forbes. Biostomp: a biocontrol system for embodied performance using mechanomyography. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2017.

[10] S. Fels and F. Vogt. Tooka: explorations of two person instruments. In *Proceedings of the International Conference on New interfaces for Musical Expression*, 2002.

[11] J. Françoise, S. Fdili Alaoui, T. Schiphorst, and F. Bevilacqua. Vocalizing dance movement for interactive sonification of laban effort factors. In *Proceedings of the Conference on Designing Interactive Systems*, pages 1079–1082. ACM, 2014.

[12] R. I. Godøy. Sonic object cognition. In *Springer Handbook of Systematic Musicology*, pages 761–777. Springer, 2018.

[13] A. Grinberg. *Touch Divided: artistic research in duo piano performance*. PhD thesis, The University of Queensland, 2017.

[14] T. Großhauser, B. Bläsing, C. Spieth, and T. Hermann. Wearable sensor-based real-time sonification of motion and foot pressure in dance teaching and training. *Journal of the Audio Engineering Society*, 60(7/8):580–589, 2012.

[15] T. Hermann, A. Hunt, and J. G. Neuhoff. *The sonification handbook*. Logos Verlag, Berlin, 2011.

[16] A. R. Jensenius. Exploring music-related micromotion. In C. Wöllner, editor, *Body, Sound and Space in Music and Beyond: Multimodal Explorations*, pages 29–48. Routledge, Oxon, 2017.

[17] A. R. Jensenius. Sonic Microinteraction in "the Air". In M. Lesaffre, P.-J. Maes, and M. Leman, editors, *The Routledge Companion to Embodied Music Interaction*, pages 431–439. Routledge, New York, 2017.

[18] A. R. Jensenius and K. A. V. Bjerkestrand. Exploring micromovements with motion capture and sonification. In *International Conference on Arts and Technology*, pages 100–107. Springer, 2011.

[19] S. Jordà. Multi-user instruments: models, examples and promises. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 23–26, 2005.

[20] S. Jordà. On stage: the reactable and other musical tangibles go real. *International Journal of Arts and Technology*, 1(3-4):268–287, 2008.

[21] M. Kaltenbrunner, S. Jorda, G. Geiger, and M. Alonso. The reactable*: A collaborative musical instrument. In *15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 406–411, 2006.

[22] S. Katan. Using interactive machine learning to sonify visually impaired dancers' movement. In *Proceedings of the 3rd International Symposium on Movement and Computing*.

[23] S. Landry and M. Jeon. Participatory design research methodologies: A case study in dancer sonification. In *The International Conference on Auditory Display*, Pennsylvania State University, 2017.

[24] T. G. Leighton. *The Acoustic Bubble*. Academic Press, London, 1994.

[25] D. Lepkoff. What is Release Technique? *Movement Research Performance Journal*, 19, 1999.

[26] C. P. Martin, A. R. Jensenius, and J. Torresen. Composing an ensemble standstill work for Myo and Bela. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2018.

[27] S. D. Monache, P. Polotti, and D. Rocchesso. A toolkit for explorations in sonic interaction design. In *Proceedings of the Audio Mostly Conference*, 2010.

[28] L. A. Naveda and M. Leman. Sonification of samba dance using periodic pattern analysis. In *Artech08*, pages 16–26. Portuguese Católica University, 2008.

[29] R. Niewiadomski, M. Mancini, A. Cera, S. Piana, C. Canepa, and A. Camurri. Does embodied training improve the recognition of mid-level expressive movement qualities sonification? *Journal on Multimodal User Interfaces*, pages 1–13, 2018.

[30] M. Ortiz. *Towards an idiomatic compositional Language for biosignal Interfaces*. PhD thesis, Queen's University Belfast, 2010.

[31] G. Paine. Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, 14(2):142–155, 2009.

[32] G. Paine. Interaction as material: The techno-somatic dimension. *Organised Sound*, 20(1):82–89, 2015.

[33] A. Phinyomark, P. Phukpattaranont, and C. Limsakul. Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8):7420–7431, 2012.

[34] A. C. E. Purser. 'being in your body' and 'being in the moment': the dancing body-subject and inhabited transcendence. *Journal of the Philosophy of Sport*, 45(1):37–52, 2018.

[35] D. Rocchesso and F. Fontana. *The Sounding Object*. Mondo estremo, Firenze, 2003.

[36] J. C. Schacher. Motion to gesture to sound: Mapping for interactive dance. In *Proceedings of The International Conference on New interfaces for Musical Expression*, pages 250–254, 2010.

[37] S. Serafin. *The sound of friction: Real time models, playability and musical applications*. PhD thesis, Stanford University, 2004.

[38] W. Siegel and J. Jacobsen. The challenges of interactive dance: An overview and case study. *Computer Music Journal*, 22(4):29–43, 1998.

[39] J. O. Smith III. *Physical Audio Signal Processing*. `http:`http://ccrma.stanford.edu/ jos/pasp/`//-ccrma.stanford.edu/~jos/pasp/`. online book, 2010 edition.

[40] J. O. Smith III. Viewpoints on the history of digital synthesis. In *Proceedings of the International Computer Music Conference*, 1991.

[41] K. Stockhausen. Four criteria of electronic music. *Stockhausen on Music: Lectures and Interviews. New York: Marion Boyars*, 1989.

[42] V. Straebel and W. Thoben. Alvin lucier's music for solo performer: experimental music beyond sonification. *Organised Sound*, 19(1):17–29, 2014.

[43] A. Tanaka. Intention, effort, and restraint: The EMG in musical performance. *Leonardo*, 48(3):298–299, 2015.

[44] K. Tatar and P. Pasquier. Musical agents: A typology and state of the art towards musical metacreation. *Journal of New Music Research*, pages 1–50, 2018.

[45] D. Trueman and R. Dubois. Percolate: a collection of synthesis, signal processing, and video objects (with source-code toolkit) for max/msp/nato v. 1.0 b3. *Computer Music Centre: Columbia University*, 2001.

Paper II

Gestures in Ensemble Performance

II

Alexander Refsum Jensenius, Çağrı Erdem

To be published in *Together in Music: Participation, Coordination, and Creativity in Ensembles*, pp. 109–118.
January 2022

Paper III

RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation

Çağrı Erdem, Alexander Refsum Jensenius

Published in *International Conference on New Interfaces for Musical Expression1*, pp. 477–482.
July 2020

III

RAW: Exploring Control Structures for Muscle-based Interaction in Collective Improvisation

Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
cagri.erdem@imv.uio.no

Alexander Refsum Jensenius
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
a.r.jensenius@imv.uio.no

ABSTRACT

This paper describes the ongoing process of developing *RAW*, a collaborative body–machine instrument that relies on 'sculpting' the sonification of raw EMG signals. The instrument is built around two Myo armbands located on the forearms of the performer. These are used to investigate muscle contraction, which is again used as the basis for the sonic interaction design. Using a practice-based approach, the aim is to explore the musical aesthetics of naturally occurring bioelectric signals. We are particularly interested in exploring the differences between processing at audio rate versus control rate, and how the level of detail in the signal—and the complexity of the mappings—influence the experience of control in the instrument. This is exemplified through reflections on four concerts in which *RAW* has been used in different types of collective improvisation.

Author Keywords

Improvisation, EMG, biosignals, sonification, mapping, ensemble, co-performance

CCS Concepts

•**Applied computing** → **Sound and music computing;** Performing arts; •**Human-centered computing** → *User centered design;*

1. INTRODUCTION

Over the last decades, we have seen a growing number of artist-researchers use the human body as part of their musical instrument. Rapid technological advancements now allow for capturing 'overt' information about human bodily processes (motion tracking), as well as measuring 'covert' processes (physiological measurements). As opposed to most traditional musical instruments, these new instruments are often 'touchless,' allowing for the creation of sonic interaction in the 'air' [13].

One challenge with playing such air instruments, is that the performance may bridge over to the aesthetics of theater acting and dance. We will leave that problem aside here, and focus on the types of air performance that is clearly situated within a context of music. Still there are several conceptual and practical challenges in how such instruments should be created. For example, how does one handle different spatiotemporal levels when not being restricted to a physical instrument? How does the design choices related to the spatiotemporal properties influence the perception of the performance? And, the question that is the main focus of this paper: how is it possible to create an 'air instrument' that can effectively be used in the context of group improvisation? From what we have seen, the majority of instruments developed for 'air performance' have focused on solo performance and/or a particular composition. But how is it possible to create a more open-ended instrument that can be used in collaborative musicking?

Figure 1: The second performance of *RAW* at the Web Audio Conference 2019 in Trondheim.

In this paper we report on the ongoing process of exploring improvisational concepts within the construction of *RAW*. Its building blocks range from the raw electromyographic (EMG) signals at audio rate, to the algorithmic approaches at control rate. Particular attention has been devoted to also interacting with other ensemble members via data interaction. After discussing the implementation, we present our subjective evaluation of using *RAW* in ecological conditions, and how that has informed the design and performance strategies.

2. BACKGROUND

2.1 Collective improvisation

In improvised music, freedom does not arise just from the notion of surprise and high complexity, but from doing so in appropriate and moderate ways [4]. Sawyer describes this as the "collaboratively emergent" nature of the group creativity, which enables something novel and coherent to occur [30]. Collective improvisation can therefore be seen as a case in which the creative agency is equally distributed among the ensemble members, which result in strict yet ever-changing constraints on an individual's creativity [15].

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

2.2 Interaction dynamics

Borgo argues that the musical development of an improvising ensemble is unpredictable, and is based on the collective dynamics and decision-making of the group [4]. This can be thought of as similar to theories of nonlinear dynamical systems, in which complex neurobiological systems adapt and change their states through self-organization [7]. For example, imagine a double pendulum and how small changes in the initial angle, mass and speed conditions of the pendulum would influence the overall motion. This can be seen as similar to the interaction dynamics of 'forces' within an ensemble. Sawyer suggests that there is always an emergent intentionality in co-creation, based on a moment-to-moment contingency [30]. The result is that any action can be altered by the subsequent energy influxes from other agencies, be that of a performer, the audience, or a machine.

2.3 Mapping: control vs uncontrol

Mapping can be described as the conveying and perceiving of physical energy, and is in many ways at the core of an instrument [12, 39]. While many mappings may be seen as one-directional and deterministic, there are also mapping strategies that are based on exploring the boundary between control and 'uncontrol.' The latter can be seen as a type of mapping in which the performer has less direct influence over the instrument. In Snyder's *The Birl*, for example, the artificial neural network (ANN) that is responsible for the mapping, outputs a 'wrong' value whenever the input exceeds a certain threshold [34]. Similarly, Kiefer emphasizes unpredictability as a more expressive (un)control paradigm using nonlinear Echo State Networks (ESNs) [14]. Schacher and colleagues also aim at the breakpoints of the machine learning algorithms to inject a creative unpredictability in their instrument called *Double Vortex* [31]. Berdahl and colleagues focus on "razor-thin edge of chaos" sound synthesis techniques [2], while Mudd et al also explore the potential of nonlinear dynamical processes for the development of new creative digital technologies [22].

2.4 From biofeedback to biocontrol

Alvin Lucier's pioneering work, *Music for Solo Performer* (1964) for "enormously amplified brainwaves" [36], was the first musical piece to explore the complex and emergent behaviors of the human physiological system. Ironically, it relied on the performer's passive states, which may be thought of as a "biofeedback" paradigm [24]. Starting in the 1990s, we have seen a paradigm shift towards "biocontrol." This paradigm was first staged by Atau Tanaka's *Kagami*, featuring *The BioMuse* [19]. Later we have seen a further shift from control to a form of co-adaptation and configuration between the body and the system [38], such as in Tanaka's *Myogram* [37] and Donnarumma's *Ominous* [9]. While most of the experimentation has been done by solo performers, there are also a few examples of ensemble works using biosystems, including The Biomuse Trio [20] and Van Nort's collaborative sound-painting [23].

3. CONCEPTUAL DESIGN

Collective improvisation implies the exploration of relationships between players [1]. This may be based on balancing between "coherence" and "inventiveness" [29], or complexity vs comprehensibility, control vs uncontrol, and constancy vs unpredictability [3]. When setting out to develop *RAW*, one of our ideas was to rely on the EMG signals coming directly from the sensors. In their "uncooked" state, these signals are inherently noisy. They are also both controllable and uncontrollable at the same time. Since we are working with the raw sensor signals, we get a signal that is highly responsive, yet at the same time quite noisy.

There are two core ideas of *RAW*:

1. Explore the naturally occurring bioelectric signals at audio rate, and use these signals as the basis for the sound synthesis.

2. Build a set of control structures that range from being limited and constrained to highly open and surprising.

Together these two approaches allow for leveraging the full dynamics of the body motion at different spatiotemporal levels. It also makes it possible to exploit the stochastic and non-stationary characteristics of EMG signals [26], at an audible level. Conceptually, this is based on explorations of unconscious processing happening while playing [6]. This is also in line with the 'post-biocontrol' paradigm mentioned in 2.4, and will allow for using the system in relevant musical idea spaces of improvisation.

The development of *RAW* has been done using a practice-based approach and iterative design methodology. That is, once we had a working prototype, we started to use the instrument in live performances with different ensembles, each of which were evaluated and the feedback used to inform the continued development process.

4. IMPLEMENTATION

4.1 Hardware setup

The hardware setup of *Raw* includes:

- two Myo armbands placed on the forearms
- a laptop running a Python script for sensor data acquisition and a Max/MSP patch for sound sculpting
- a sound interface for audio I/O
- an iPad running the *Mira* app

The signal flow is sketched in Figure 2, and we will in the following go through each of the core components in detail.

4.2 EMG Data acquisition

EMG signals represent the electrical activity produced by muscles [26]. Each of the Myo armbands is equipped with 8 EMG sensors that are sampled at a rate of 200 Hz. Based on knowledge from hand-gesture recognition models [27], we decided to use the 4^{th} and 8^{th} Myo sensors. These correspond to the *extensor carpi radialis longus* and *flexor carpi radialis* muscles, respectively.

Since we have experienced a lot of problems in the past with Bluetooth-based devices, and particularly when using multiple devices at the same time, we decided to develop our own data acquisition solution.[1] This is a custom Python script based on Martin's *myo-to-osc* [21]. Here we implemented low-latency support for multiple Myo armbands, each connecting to the computer via separate Bluetooth Low Energy (BLE) adapters. This was important to overcome bandwidth limitations and data dropouts. The script can also be used to store data from the devices together with audio. This is useful to document and evaluate the latency and jitter of the data stream, and also for further analysis and model building. The script runs as multiple processes: data acquisition from the 1^{st} and 2^{nd} armbands, and audio recording using *PyAudio* [25], respectively.

[1] `https://github.com/chaosprint/dual-myo-recorder`

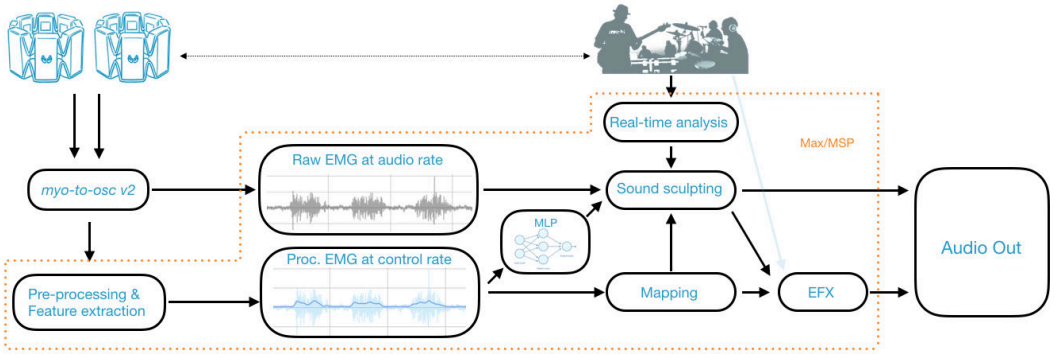


Figure 2: Simplified signal flow diagram for the performance of *RAW*.

4.3 Sound sculpting

The sound generating part of *RAW* relies on 'sculpting' raw EMG signals at audio rate. The incoming raw signals (2 channels per arm) are first normalized, and then written recurrently into buffers every 50 samples (250ms). This is below the constraint of a 300-ms acceptable delay [10], and provides four dynamic wavetables that are continuously updated. Then, the wavetables are brought to an audible range and their frequency spectra are controlled by time-scaled sawtooth signals. Finally, the buffers are 'sculpted' using direct audification of the raw EMG signals via processed control signals mapped to low-level MSP operators.

The second sound module uses a database of recorded percussive sounds of 1–5-s duration. As opposed to the sustained signal quality of the sonified muscle signals, this module provides a pointillistic way of wave-shaping. In addition, we also use the aforementioned wavetable strategy for external audio input reserved for other ensemble members. This allows for both sound sculpting and live processing throughout the performance.

4.4 Control signals

The signal coming from an EMG sensor is fairly complex, due to its stochastic and noisy nature. This is interesting at audio rate, but poses more challenges when used to create meaningful control signals. The first part of the signal chain is based on a fourth-order Butterworth filter with bandpass at 20–200 Hz. Second, we apply feature extractors to reduce the dimension of the discrete signals into a better representation. Here we take the root mean square (RMS) of the signal to represent the overall energy trend.

RMS works well for extracting larger-scale events from the EMG signal. However, one might consider alternative features for a better responsiveness to agility in motion. For that purpose, we relied on nonlinear Bayesian filtering (using the *pipo.bayesfilter* Max external object) as it provides significant advantages for the amplitude estimation of 'sudden changes' [11], as opposed to estimators such as the RMS that trims 'bumpy' information for a better trend.

As we do not use a physical interface, triggering sonic events in a more time-sensitive manner can become a challenging task. To tackle this issue, a relevant strategy is to detect the onsets, or, in other words, to determine the period of muscle activation based on the amplitude of the EMG signal. Among a range of methods, we relied on the Teager-Kaiser Energy (TKE) operation [17] for the muscle onset detection. We included TKE extractor in the Python script to process the signals in time domain as $y(n) = x^2(n) - x(n-1)x(n+1)$.

4.5 Attractor states

In *RAW* we program the compositional 'motives' based on *attractors*. This is inspired by the fields of dynamical systems, in which an attractor represents a set of points in space that evolve using differential equations. These equations draw identifiable trajectories in the *phase space* [18], illustrating broad outlines of complex behavior.

Our implementation is based on a Support Vector Machine (SVM) classifier that recognizes the pinch grips of the performer. These are then drawn as a new set of points on the orbit that is mapped to sound synthesis parameters. The non-periodic and unstable behavior of these attractors trigger seemingly random spectro-temporal events. Yet, the trajectories accumulate to a final shape that looks 'attracted' to the compositional motif, such as in using a pre-written chord progression.

In addition to the SVM classifier, we employ various random processes in the mapping structure, based on Brownian noise. Random values are preferred for the exponential base of scaling curves, as well as for wave-shaping and amplitude modulation. This is to create a more uncertainty than what is typically achieved with linear mapping structures.

4.6 Machine learning

The system uses supervised Multi-Layer Perceptron (MLP) algorithms for regression, using the *ml.** library for Max [5]. Each artificial neural network (ANN) is set up with three hidden layers, relying on bipolar sigmoid activation functions to map the 8-dimensional EMG data to a 2D-point on an XY plane. The ANNs are trained on a dataset consisting of hand waving and a detour on the plane. In other words, the start (x=0, y=0) and endpoints (x=127, y=0) are constant, while the trajectory is nonlinear. This can be thought of as a 'gamified' strategy. Imagine having a 'ball' (point) on each hand, sharing the same plane, in which the goal is to make the two balls intersect to successfully trigger and/or adjust the events. The performer is then required to have a clear imagery of the plane to have complete control on the generation of events. In most cases the performer will make 'mistakes,' willingly or unwillingly, which will lead to unexpected events.

4.7 Ensemble interaction

An important feature of *RAW* is the implementation of strategies that allow for direct interaction with an ensemble.

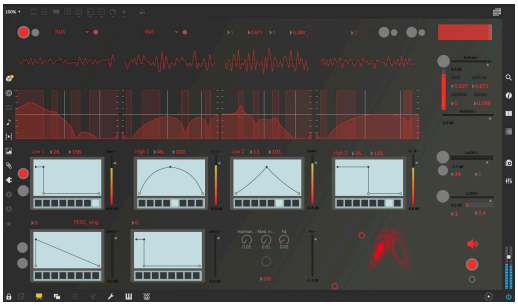


Figure 3: The graphical user interface (GUI) of *RAW*, designed in Max/MSP.

One of the ways to achieve this is through a real-time audio analysis module that is programmed to interact with any kind of audio input. This module is implemented using a chain consisting of: an envelope follower based on a median filter; a tempo tracker using the btrack~ object [35]; an onset detector using the spectral flux for controlling temporal events, such as stutter, tremolo and delay; and *pipo*~ plugins from the Ircam MuBu library [32] for spectral analysis.

4.8 Outboard

The last part of the sound signal flow in *RAW* is a small set of effects. This includes: a simple reverberator based on the Schroeder model [33]; delay lines based on comb filtering effect (*comb*~); and state-variable filters (*svf*~) that are driven by the time-scaled output of the interaction modules.

5. PERFORMANCES

RAW has been used in four public performances to date,[2] each with a different ensemble. We will in the following discuss how the different performances have shaped the performance strategies and the development of the instrument.

5.1 Ensemble 1: Trio with live coding, voice & body resonators

The premier performance of *RAW* took place in a cultural center in Oslo, Norway. The ensemble featured Tejaswinee Kelkar, performing with kitchen utensils actuated through her voice, and Qichao Lan using a live coding environment, Quaver Series ([16]), designed by himself. The 20-minute set was structured as short solo acts of each musician, followed by a collective improvisation.

The *RAW* solo started with a short interlude with the vocalist. The voice was fed into the instrument and processed using modest time-stretching, controlled by the performer's upper arm abduction and wrist flexion/extension. Here we observed how the (spoken) voice influenced the body motion of the performer in a particular way, which was largely based on sustained motion with occasional impulses. The muscle tension was generally low, and its fluctuations were slightly perceivable. This interplay set an example for how a combination of creative interactions and emerging constraints enable an experience of *flow* [8].

The collaborative improvisation part evolved into the use of rhythmic structures. Here we observed two distinct layers: a set of pulse-based rhythms, and a set of discontinuous dynamic (accelerating vs decelerating) rhythms with intermittent textures. Drawing on Grisey's continuum of rhythm (as elaborated in [28]), each of these rhythmic structures represented two extremes: 'Order' (predictability) on one end, and 'disorder' (unpredictability) on the other. Finally, the higher complexity of rhythmic structures steered *RAW* towards a higher rhythmic complexity as well, quite different from its smooth and sustained trend in the solo section.

[2]Videos available at `http://bit.ly/raw_videos`

5.2 Ensemble 2: Quintet with live coding, shared electric guitar & laptop, voice & laptop

This performance (Figure 1) was part of the Web Audio Conference 2019 (WAC) in Trondheim, Norway. It also featured live coder Steven Yi, together with Ariane Stolfi on live processed voice, and Luis Arandas and Michel Buffa who shared a guitar and a laptop. In live coding, the musician writes code on the computer to generate sounds. The striking aspect of this performance style is that it heavily relies on the machine clock rather than human bodily rhythms. So in a collaborative performance, the human performers naturally tend to align with how the live coder structures the (machine) time.

The first salient feature of this collaborative performance was the gentle pulses coming from a performer on a Csound-based live coding environment. While the live-coded sound shapes were more 'vertical,' the rest of the ensemble played more sustained sonic patterns. The first half of the performance demonstrated a mellow and ambient musical structure, along with short-lived dynamic articulations.

Borgo speaks about two types of transitions in free improvisation: small-scale transitions that occur dynamically between different parties within the ensemble, and larger-scale transitions that happen through complete synchrony and flow [4]. In this performance we observed one larger-scale transition between the two halves of the performance. There was also a dynamic interplay happening between *RAW* and the voice, while the guitar maintained the ambient layer along with live coded pulses. It was interesting to observe, once again, a naturally occurring musical coupling between processed voice and a muscle-based instrument, which should be further investigated. Finally, in this performance *RAW* relied on control structures of low complexity, which showcased the potential of using gentle, sustained, body motion in muscle-based performance.

5.3 Ensemble 3: Duo with gestural controller

This duo performance was part of a special event for gestural interaction, which took place in a nightclub in Istanbul, Turkey (Figure 4). The ensemble featured *RAW* together with *Armonic*, a gestural control system based on inertial measurement units (IMUs) and capacitive sensors. *Armonic* specializes in a gestural live sampling technique, with a particular focus on precision and control. Görkem Arıkan, the inventor of *Armonic*, draws an analogy between his performance style and 'puppetry:' controlling "sounds through 'invisible' ropes prolonging from [his] hands."

This was a quite different performance, in that both performers played on 'air instruments.' Thus, even though both of the instruments were untraditional, they shared some similar affordances. This, combined with the coziness of a small club stage, allowed both performers to develop an interpersonal language beyond their normal strategies for action–sound mappings. This was experienced as being similar to how dancers often do contact improvisation (CI), in which the emphasis is put on inter-corporeal experimentation, curiosity, and self-surprise [15].

This relatively short ($10'$) performance demonstrated rapidly changing idea spaces, and several larger-scale transitions. The overall trend of rhythmic structures alternated between the two extremes of the before-mentioned Grisey's continuum (from *smooth* to *random*), which also resulted in an

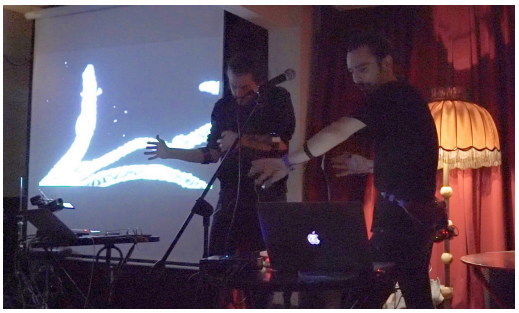


Figure 4: Duo performance by Çağrı Erdem and Görkem Arıkan in Istanbul 2019. The improvised performance featured two 'air instruments': *RAW* and *Armonic* (Photo: Mehmet Ömur)

energy trend that varied almost idiosyncratically. In the case of *RAW*, this led to a dynamic interplay based on an extensive use of control structures of high complexity, particularly the ANN-based 'gamified' strategy.

5.4 Ensemble 4: Duo with drums

This duo performance featured *RAW* together with drummer Onur Başkurt, and took place in a jazz club in Istanbul, Turkey, as part of a larger event that hosted several improvisation ensembles. The drummer used a small drum set, and it was not equipped with microphones, except for an overhead ribbon microphone that was used to capture audio for *RAW*.

The performance was based on a rough sketch of playing three sections in A-B-A′ form. The A section was focused on exploring the full dynamic potential of a muscle-based instrument. *RAW* relied on sound-sculpting the muscle signals, using both extremes of the dynamic range actively. This first part of the performance was completely led by muscle contraction effort processed at audio rate. The drummer tried to carefully follow the dynamic fluctuations, using accelerating and decelerating rhythms. This echoed how Grisey indicates the intrinsic relationship between tension and discontinuous rhythmical dynamics.

The B section opened with a short drum solo interlude. *RAW* eventually joined in with chopped-up (iterative) samples. In this section, we observed how muscles are intrinsic to small-scale body motion that is hardly perceivable.

The A′ section was mostly a recap of A, with an additional closing as it was the end of the performance. All in all, *RAW*'s strict control over the dynamic shape, combined with unexpected timbral outcomes, led to an interesting combination of controllability and surprise.

6. CONCLUSIONS

The central ideas of *RAW* were to explore the raw EMG signals at audio rate, and to build a set of control-level mapping structures. Already from the first prototype and performance, this worked quite well. Subsequent performances were important for further exploration, evaluation, and modification of the system.

One important finding from the development, is that there is a huge difference in the mean amplitude of muscle signals at rest versus during performance. Such changes in psychophysiological conditions are important to bear in mind when developing a muscle-based instrument, and are not possible to test without carrying out real-world performances.

Another finding is that of the importance of a certain level of causality between action and sound. This became particularly evident in Ensemble 3, in which both performers played with 'air instruments.' Here both performers used full-range sound spectra distributed through the same sound system. This caused problems of masking and lack of spatialization.

All in all, we find *RAW* to be a well-functioning instrument, and it has proved to be stable in real-world performance contexts. Still there are numerous things to improve in future iterations:

- Action–Sound Causality: Even though a 'blind' exploration of (musical) gestures may be exciting at first, performing with different ensembles ascertained the necessity of a certain level of causality between action and sound, hence the possibility of repeatable playing technique. Since we are working at a level of muscle-control, future developments will include explorations of fine motor patterns. Through this we aim to improve the mapping structures and interactive affordances of the instrument.
- Interaction: Unpredictable processes work well in small-scale transitions, since these moments allow for 'debate' between performers. Moreover, such processes showed that simple mappings can be engaging and serendipitous. However, a whole-group synchrony is crucial for transitions of larger musical idea spaces. This is where traditional instruments allow for a superior responsiveness and causality than most 'air instrument' designs. To this end, we will focus on better machine listening and real-time interaction strategies.
- Rhythm tracking: Performing with a drum set in Ensemble 4 revealed the necessity for implementing a better non-periodic rhythm-tracking, and developing 'riff-based' playing techniques.
- Spatiotemporality: Each of the co-performing instruments have had unique spatiotemporal characteristics, which combined with the spatial range, metabolism and biomechanics of the human body, have led to many interesting audiovisual moments. In live coding, for example, you sit, and write and rewrite text. When playing a drum set, you also sit, surrounded by several physical objects of different sizes, shapes and materials. A muscle-based 'air instrument' is not bound to the same type of physical space, but this still leads to many questions about how space should be used, how time should be structured, and how to interact audiovisually with the other performer(s).

These conceptual and practical challenges will be addressed in our future developments of muscle-based performance.

7. ACKNOWLEDGMENTS

This work was partially supported by the Research Council of Norway (project 262762) and NordForsk (project 86892).

8. REFERENCES

[1] D. Bailey. *Improvisation: Its Nature and Practice in Music*. Da Capo Press. Originally published in 1992, New York, NY, 1993.

[2] E. Berdahl, E. Sheffield, A. Pfalz, and A. T. Marasco. Widening the razor-thin edge of chaos into a musical highway: Connecting chaotic maps to digital waveguides. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, Blacksburg, VA, 2018.

[3] D. Borgo. Negotiating freedom: Values and practices in contemporary improvised music. *Black Music Research Journal*, pages 165–188, 2002.

[4] D. Borgo and J. Goguen. Rivers of consciousness: The nonlinear dynamics of free jazz. In *Jazz research proceedings yearbook*, volume 25, Long Beach, CA, 2005.

[5] J. Bullock and A. Momeni. Ml. lib: robust, cross-platform, open-source machine learning for max and pure data. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, Baton Rouge, LA, 2015.

[6] D. Chi, M. Costa, L. Zhao, and N. Badler. The emote model for effort and shape. In *Proc. 27th annual Conf. on Computer graphics and interactive techniques*, New York, NY, 2000.

[7] J. Y. Chow, K. Davids, R. Hristovski, D. Araújo, and P. Passos. Nonlinear pedagogy: Learning design for self-organizing neurobiological systems. *New Ideas in Psychology*, 29(2):189–200, 2011.

[8] M. Csikszentmihalyi. *Creativity: Flow and The Psychology of Discovery and Invention*. Harper Collins, New York, NY, 1996.

[9] M. Donnarumma. Ominous: Playfulness and emergence in a performance for biophysical music. *Body, Space & Technology*, 14, 2015.

[10] K. Englehart and B. Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *IEEE transactions on biomedical engineering*, 2003.

[11] D. Hofmann, N. Jiang, I. Vujaklija, and D. Farina. Bayesian filtering of surface emg for accurate simultaneous and proportional prosthetic control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(12):1333–1341, 2015.

[12] A. Hunt and M. M. Wanderley. Mapping performer parameters to synthesis engines. *Organised sound*, 7(2):97–108, 2002.

[13] A. R. Jensenius. Sonic Microinteraction in "the Air". In M. Lesaffre, P.-J. Maes, and M. Leman, editors, *The Routledge Companion to Embodied Music Interaction*, pages 431–439. Routledge, New York, 2017.

[14] C. Kiefer. Musical instrument mapping design with echo state networks. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, London, UK, 2014.

[15] M. Kimmel, D. Hristova, and K. Kussmaul. Sources of embodied creativity: interactivity and ideation in contact improvisation. *Behavioral Sciences*, 2018.

[16] Q. Lan and A. R. Jensenius. Quaverseries: A live coding environment for music performance using web technologies. In *Proc. Int. Web Audio Conf.*, Trondheim, Norway.

[17] X. Li, P. Zhou, and A. S. Aruin. Teager–kaiser energy operation of surface emg improves muscle activity onset detection. 2007.

[18] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 1963.

[19] H. S. Lusted and R. B. Knapp. Biomuse: Musical performance generated by human bioelectric signals. *The Journal of the Acoustical Society of America*, 84(S1):S179–S179, 1988.

[20] E. Lyon, R. B. Knapp, and G. Ouzounian. Compositional and performance mapping in computer chamber music: A case study. *Computer Music Journal*, 38(3):64–75, 2014.

[21] C. P. Martin, A. R. Jensenius, and J. Torresen. Composing an ensemble standstill work for myo and bela. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, Blacksburg, VA, 2018.

[22] T. Mudd, S. Holland, and P. Mulholland. Nonlinear dynamical processes in musical interactions: Investigating the role of nonlinear dynamics in supporting surprise and exploration in interactions with digital musical instruments. *International Journal of Human-Computer Studies*, 2019.

[23] D. V. Nort. Conducting the in-between: improvisation and intersubjective engagement in soundpainted electro-acoustic ensemble performance. *Digital Creativity*, 29(1):68–81, 2018.

[24] M. Ortiz-Perez, N. Coghlan, J. Jaimovich, and R. B. Knapp. Biosignal-driven art: Beyond biofeedback. *Ideas Sonica/Sonic Ideas*, 3(2), 2011.

[25] H. Pham. Pyaudio: Portaudio v19 python bindings. *URL: https://people. csail. mit. edu/hubert/pyaudio*, 2006.

[26] A. Phinyomark, E. Campbell, and E. Scheme. Surface electromyography (emg) signal processing, classification, and practical considerations. In *Biomedical Signal Processing*. Springer, 2020.

[27] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. Application of wavelet analysis in emg feature extraction for pattern classification. *Measurement Science Review*, 2011.

[28] C. Roads. *Composing electronic music: a new aesthetic*. Oxford, UK, 2015.

[29] R. K. Sawyer. Learning music from collaboration. *International Journal of Educational Research*, 47(1):50–59, 2008.

[30] R. K. Sawyer and S. DeZutter. Distributed creativity: How collective creations emerge from collaboration. *Psychology of aesthetics, creativity, and the arts*, 3(2):81, 2009.

[31] J. C. Schacher, C. Miyama, and D. Bisig. Gestural electronic music using machine learning as generative device. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, Baton Rouge, LA, 2015.

[32] N. Schnell, A. Röbel, D. Schwarz, G. Peeters, R. Borghesi, et al. Mubu and friends–assembling tools for content based real-time interactive audio processing in max/msp. In *ICMC*, Montreal, Quebec, Canada, 2009.

[33] J. O. Smith III. *Physical Audio Signal Processing*. `http:`http://ccrma.stanford.edu/ jos/pasp/`//-ccrma.stanford.edu/~jos/pasp/`. online book, 2010 edition.

[34] J. Snyder and D. Ryan. The birl: An electronic wind instrument based on an artificial neural network parameter mapping structure. In *Proc. Int. Conf. on New Interfaces for Musical Expression*, London, UK, 2014.

[35] A. M. Stark, M. E. Davies, and M. D. Plumbley. Real-time beat-synchronous analysis of musical audio. In *Proc. of the Int. Conf. on Digital Audio Effects, Como, Italy*, Como, Italy, 2009.

[36] V. Straebel and W. Thoben. Alvin lucier's music for solo performer: experimental music beyond sonification. *Organised Sound*, 19(1):17–29, 2014.

[37] A. Tanaka. Myogram, metagesture music cd, 2017.

[38] A. Tanaka and M. Donnarumma. The body as musical instrument. *The Oxford Handbook of Music and the Body*, 1, 2018.

[39] M. M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proc. IEEE*, 2004.

Paper IV

Exploring relationships between effort, motion, and sound in new musical instruments

Çağrı Erdem, Qichao Lan, Alexander Refsum Jensenius

Published in *Human Technology*, pp. 310–347.
November 2020

IV

HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 16(3), November 2020, 310–347

EXPLORING RELATIONSHIPS BETWEEN EFFORT, MOTION, AND SOUND IN NEW MUSICAL INSTRUMENTS

Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Qichao Lan
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Alexander Refsum Jensenius
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
University of Oslo
Norway

Abstract: *We investigated how the action–sound relationships found in electric guitar performance can be used in the design of new instruments. Thirty-one trained guitarists performed a set of basic sound-producing actions (impulsive, sustained, and iterative) and free improvisations on an electric guitar. We performed a statistical analysis of the muscle activation data (EMG) and audio recordings from the experiment. Then we trained a long short-term memory network with nine different configurations to map EMG signal to sound. We found that the preliminary models were able to predict audio energy features of free improvisations on the guitar, based on the dataset of raw EMG from the basic sound-producing actions. The results provide evidence of similarities between body motion and sound in music performance, compatible with embodied music cognition theories. They also show the potential of using machine learning on recorded performance data in the design of new musical instruments.*

Keywords: *EMG, music, machine learning, musical instrument, motion, effort, guitar, embodied.*

DOI: https://doi.org/10.17011/ht/urn.202011256767

INTRODUCTION

What are the relationships between action and sound in instrumental performance, and how can such relationships be used to create new instrumental paradigms? These two questions inspired the experiments presented in this paper. Our research is based upon two basic premises: It is possible to find relationships between the continuous, temporal shape of an action and its resultant sound and that embodied knowledge of an existing instrument can be translated into a new performative context with different instrument. Thus, we are interested in exploring whether it is possible to create mappings in new instruments based on measured actions on and sounds from an existing instrument. It is common to create such action–sound mappings based on overt motion features. However, in our study, we were interested primarily in exploring whether covert muscle signals can be used for new musical instruments.

Embodied Knowledge

The body's role in the experience of sound and music is central to the embodied music cognition paradigm (Leman, 2008). Several studies have explored the embodiment of musical experiences by investigating how musicians and nonmusicians transduce what they perceive as musical features into body motion. Sound-tracing is one such experimental paradigm that has been used to study how people spontaneously follow salient features in music (Kelkar, 2019; Kozak, Nymoen, & Godøy, 2012; Nymoen, Caramiaux, Kozak, & Torresen, 2011). Sound mimicry is a similar approach, based on examining how sound-producing actions can be imitated "in the air," that is, without a physical interface (Godøy, 2006; Godøy, Haga, & Jensenius, 2005; Valles, Martínez, Ordás, & Pissinis, 2018). Several other studies have aimed at identifying musical mapping strategies, drawing on concepts of embodied music cognition as a starting point (e.g., Caramiaux, Bevilacqua, Zamborlin, & Schnell, 2009; Françoise, 2015; Maes, Leman, Lesaffre, Demey, & Moelants, 2010; Tanaka, Donato, Zbyszynski, & Roks, 2019; Visi, Coorevits, Schramm, & Miranda, 2017).

In this study, we took bodily imitation as the starting point for the creation of action–sound mappings. The idea was to transfer the acquired skills of playing traditional instruments to a new context. Here the term traditional refers to the recognizability of performance skills, what Smalley (1997) explained as an intuitive knowledge of action–sound causalities in traditional sound-making. The idea was to exploit such proprioceptive relationships between musician and instrument (Paine, 2009). The premise is that skill can be understood as embodied knowledge (Ingold, 2000) that leads to lower information processing at a cognitive level (Dreyfus, 2001). It also builds upon the idea that spectators can perceive and recognize skill as an embodied phenomenon (Fyans & Gurevich, 2011).

One outcome of this research was aimed at developing solutions for creating musical instruments that can be performed in the air. However, it should be clear from the start that we are not interested in making "air" versions of the guitar or any other physical instrument. Rather, our attention is devoted to reusing the embodied knowledge of one type of instrumental performance in new ways (Magnusson, 2019). The lack of a haptic and tactile experience creates a significantly different experience when playing a physical instrument as compared to a touchless air instrument. According to the "gestural agency" concept of Mendoza Garay & Thompson (2017), the instrument is as much an agent in the musical transaction as the performer:

They influence each other within a musical ecosystem. In this system, the agents' communication is multimodal. Therefore, the act of instrument playing accommodates not only the auditory, tactile, and haptic channels but also the visual, kinetic, proprioceptive, or any other kind of interactions that have a musical influence. The human agent becomes the participant that is expected to adapt; thus, any change in the environment can be seen as a creative challenge.

From Body Motion to Musical Actions

Gesture is employed frequently in the literature on music-related body motion (Cadoz & Wanderley, 2000; Gritten & King, 2011; Hatten, 2006). We understand gesture as related to the meaning-bearing aspects of performance actions. In this project, we focus not on such meaning-bearing aspects and thus will not use that term in the following discussion. Instead, we will use *motion* to describe the continuous displacement of objects in space and time, and *force* to explain what sets these objects into motion. Both motion and force are physical phenomena that can be captured and studied using various devices (see Jensenius, 2018a, for an overview of various methods for sensing music-related body motion). Hitting a guitar string is an example of what we call motion, which can be studied through motion capture data of the arm's continuous position. Muscle tension is an example of the force involved in the sound production and can be studied through electromyography (EMG).

Motion and force describe the kinematic and kinetic aspects of performance, respectively. These relate to—but are not the same as—the experienced action within a performance (Jensenius, Wanderley, Godøy, & Leman, 2010). Thus, in our research, we use *action* to describe a cognitive phenomenon that can be understood as goal-directed units of motion and/or force (Godøy, 2017). Many actions are based on visible motion, but an action also can be based solely on force. For example, some electroacoustic musical instruments are built with force-sensitive resistors that can be pressed by the performer, even without any visible motion. Hence the player's action can change drastically over time even with no or only little observable body motion.

Music-related body motion comes in various types (see Jensenius et al., 2010, for an overview). Here we primarily focus on the *sound-producing actions*. These can be subdivided into *excitation* actions, such as the right hand that excites the strings on a guitar, and *modification* actions, such as the left hand modifying the pitch. The excitation action can be divided further into the three main categories proposed by Schaeffer (2017), as sketched in Figure 1: *impulsive*, *sustained*, and *iterative*. An impulsive excitation is characterized by a fast attack and discontinuous energy transfer, while a sustained excitation has a gradual onset and continuous energy transfer. An iterative excitation is based on a series of discontinuous energy transfers.

Action–Sound Coupling and Mappings

Sound production on a traditional instrument is bound by the physical constraints of the instrument and the capabilities of human body. For example, although both are plucked instruments, a banjo, and an oud have different damping characters due to the resonant features of the instruments' bodies. The physical properties of the instruments also define their unique timbre and how they are played. Additionally, the human body has its expressive limitations. These limitations can be in

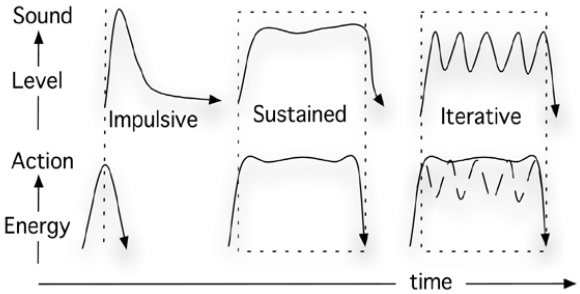


Figure 1. Illustration of the three, basic action–sound types: impulsive, sustained, and iterative (Jensenius, 2007; Used with permission).

the form of what Godøy (2018) suggested as "effort constraints," meaning "limits to endurance," which necessitate an optimization of muscle contractions (i.e., to prevent injuries). He described these limitations as also leading to "coarticulation," which results from multiple individual actions merging into larger units. All these levels of constraints are part of the transformation of biomechanical energy to sound features. We think that during the transformations in *action–sound couplings* (Jensenius, 2007), the relationships between actions and sounds are dictated by the laws of physics.

When playing a traditional instrument, one must exercise muscular exertion to abide by the instrument's physical boundaries. In the case of the guitar, this prevents the player from breaking a string due to excessive effort or not producing sound due to the lack of energy input (Tanaka, 2015a). After centuries of design, the construction of traditional instruments is no longer open to much interpretation, except for using some extended playing techniques or additional equipment. To the contrary, electroacoustic musical instruments are based on the creation of *action–sound mappings*. Here the constraints of hardware and/or software elements often are open to interpretation. In other words, the relationships between biomechanical input and the resultant sound are designed and may not correspond to each other. However, the creation of meaningful action–sound mappings is critical for how an instrument's playing and its sound are perceived (Hunt & Wanderley, 2002; Van Nort, Wanderley, & Depalle, 2014). This is often discussed as the "mapping problem" (Maes et al., 2010), which has been a central research topic in the field of new interfaces for musical expression over the last decades (Jensenius & Lyons, 2017).

New Musical Interactions

The number of artists and researchers interested in using the human body as part of their musical instrument has been growing over the last decades. Such interests often lead to the use of gestural controllers, which are types of wearable sensors or camera-based devices that allow for touchless performance, that is, a type of performance not based on touch of physical objects. As such, these instruments allow for sonic interaction in the air (Jensenius, 2017). Examples of such instruments are the Virtual Air Guitar (Karjalainen, Mäki-Patola, Kanerva, & Huovilainen, 2006), the Virtual Slide Guitar (Pakarinen, Puputti, & Välimäki, 2008), and Google's Teachable Machine, which lets users mimic guitar-playing in front of a web camera (Google, 2020).

The above-mentioned examples focus mainly on creating an air guitar. However, this is not the focus of our current research; rather, we seek to explore new ways of performing in the air. Although motion-based tracking often is employed for air instruments, we are interested specifically in measuring muscle tension through electromyography (EMG). When worn on the forearm, EMG sensors can provide muscle activation information related to the motion of hand and fingers (Kamen, 2013). EMG goes beyond measuring limb positions and provides information of the muscle articulation throughout the preparation for and execution of an action (Tanaka, 2019). The use of muscle activation data in musical performance was pioneered by Knapp & Lusted (1990) and has been practiced extensively by Tanaka (1993, 2015b). Mechanomyograms (MMGs), as a signal for muscle-based performance (Donnarumma, 2015), also have been studied.

Performing in the air introduces several conceptual and practical challenges. For example, when does a sound-producing action begin and end when no physical instrument defines the performance space? How can one handle the use of physical effort as part of that action without being restricted to a physical instrument? To address such problems, we drew on what Tanaka (2015a) suggested as an embodied interaction strategy: He replaced constraints, such as those experienced while playing a traditional instrument, with "restraints," that is, the "internalization of effort" (p. 299). Such restraints can help define a set of affordances that can replace the physical constraints found in a traditional instrument.

Even though we are interested in creating new instrument concepts, this may not necessarily require developing an entirely new action–sound repertoire. Michel Waisvisz, the creator of *The Hands* (Waisvisz, 1985), focused on maintaining the action–sound mappings of his instrument. This helped him develop and maintain a skill set over time. We propose a design strategy based on what Magnusson (2019) referred to as an "ergomimetic" structure. Here *ergon* stands for work memory and *mimesis* for imitation. Such an ergomimetic structure may help in reusing well-known interactions of a performer in a new performative context. Of course, such an approach raises some questions. For example, what types of errors and surprises emerge when a physical pipeline is replaced by software? We aim through our research to contribute to better understanding how a musician's physical skills could transfer to new air instruments.

Machine Learning

Machine learning is a set of artificial intelligence techniques for tackling tasks that are too difficult to solve through explicit programming; it is based on finding patterns in a given set of examples (Fiebrink & Caramiaux, 2016). Deep learning is a subset of machine learning, where artificial neural networks allow computers to understand complex phenomena by building a hierarchy of concepts out of simpler ones (Goodfellow, Bengio, & Courville, 2016). Machine learning has been an important component in the design of and performance with new interfaces for musical expression since the early 1990s (Lee, Freed, & Wessel, 1991). Several easy-to-use tools have been developed over the years for artists and musicians (see, e.g., Caramiaux, Montecchio, Tanaka, & Bevilacqua, 2015; Fiebrink, 2011; Martin & Torresen, 2019), and many new instruments have explored the creative potential of artificial intelligence in music and performance (Caramiaux & Donnarumma, 2020; Kiefer, 2014; Næss, 2019; Schacher, Miyama & Bisig, 2015; Tahiroğlu, Kastemaa & Koli, 2020). However, unlike the applications for generating music in the form of musical instrument digital interface (MIDI)

data (Briot, Hadjeres, & Pachet, 2020) or generating music in the wave-form domain (Purwins et al., 2019), the use of deep learning techniques for interactive music is rather rare. We see that deep learning can be particularly useful when dealing with complex muscle signals.

Research Questions

The brief theoretical discussion above has shown that a number of questions remain open regarding how musical sound is performed and perceived and how it is possible to create new empirically based sound-making strategies. Thus, in the current two-experiment study, we were interested particularly in

1. What types of muscle signals are found in electric guitar performance and how do these signals relate to the resultant sound?
2. How can we use deep learning to predict sound based on raw electromyograms?

We begin by explaining the methodological framework that has been developed for the first empirical study, followed by a presentation and discussion of the results. We then reuse some of the data from the first experiment to pursue a preliminary predictive model for action–sound mappings. We conclude with a general discussion of the findings of these two experiments.

EXPERIMENT 1: MUSCLE–SOUND RELATIONSHIPS

Methods

Research Design

This aspect of our research is based on the outcomes of an experiment with electric guitar players. Each of the guitarists performed, while wearing various sensors, a set of basic sound-producing actions as well as free improvisations. To collect the data these actions produced, we built a multimodal dataset of EMG and motion capture data; additionally, video and sound recordings of each performer were made. For this paper, we focus only on a statistical analysis of the EMG data and sound recordings from this first experiment, with a particular emphasis on similarity measures. Prior to conducting the research, we obtained ethical approval from the Norwegian Center for Research Data (NSD), Project Number 872789.

Participants

Thirty-six music students and semiprofessional musicians took part in the study. Five of the datasets turned out to be incomplete and these were excluded from further analysis. Thus, the final dataset consisted of 31 participants (30 male, 1 female, M_{age} = 27 years, SD = 7), all right-handed. All the participants had some formal training in playing the electric guitar, ranging from private lessons to university level education. The recruitment was conducted through an online invitation published on a specified web site of the University of Oslo, Norway, and announced in various communication channels targeting music students. Participation was rewarded with a gift card (valued at approximately €30).

Data Collection

The participants' muscle activity was recorded as surface EMG with two systems: consumer-grade Myo armbands and a medical-grade Delsys Trigno system. The former has a sample rate of 200 Hz, while the latter has a sample rate of 2000 Hz. Overt body motion was captured with a 12-camera Qualisys Oqus infrared optical motion capture system at a frame rate of 200 Hz. This system tracked the three-dimensional positions of reflective markers attached to each participant's upper body and the instrument. A trigger unit was used to synchronize the Qualisys and Delsys Trigno systems. Additionally, we developed a custom-built software solution to capture data from the Myo armbands in synchrony with the audio. Regular video was recorded with a Canon XF105 camera, which was synchronized with the Qualisys motion capture system. Figure 2 demonstrates the two major means for gathering data: the motion-capture configuration and the EMG system.

Procedure

Each participant was recorded individually. One recording session took 90-105 minutes. First, the participants received a brief explanation about the experiment, before they signed the consent form. Following the recording session, they completed a short survey regarding their musical background, their use of musical equipment, and their thoughts on new instruments and interactive music systems.

The participants were instructed to stand at the same marked spot in the laboratory. We asked them to perform tasks based on well-known electric guitar techniques. The hammer-on and pull-off are similar techniques that allow the performer to play multiple notes connected in a legato manner (tied together). In both techniques, the left-hand fingers hit multiple notes with a single excitation action. Hammer-on refers to bringing down another finger with sufficient force to hit a

(a) (b)

Figure 2. (a) A participant during the recording session. Motion capture cameras are visible hanging in the ceiling rig behind and on stands in front of the performer. The monitor with instructions for the performer can be seen below the front left motion capture camera. **(b)** The protocol used for placement of the EMG electrodes: Two Delsys EMG sensors were placed on each side of the arm corresponding to the extensor carpi radialis longus and flexor carpi radialis muscles, just below the Myo armbands.

neighboring note on the fretboard. Pull-off refers to moving the finger from one fret to another to modify the pitch. Bending is achieved by a finger pulling or pushing the string across the fretboard to smoothly increase the pitch. The given tasks were as follows:

- A warm-up improvisation with metronome at 70 bpm
- Task 1
  - Softly played impulsive notes B and C in 3rd and 4th octaves, respectively
  - The same task, played strongly
- Task 2
  - Softly played iterative notes
    - Single pitch (B3)
    - Double pitches (B3–C4)
  - The same task, played strongly
- Task 3
  - Softly played legato
  - The same task, played strongly
- Task 4
  - Softly played bending (semi-tone)
  - The same task, played strongly
- A free improvisation (the tone features and the use of metronome are at the participant's discretion)

We based the tasks on performing guitar-like versions of each of the three action–sound types. Tasks 1 and 4, for instance, lie somewhere in between classes considering that the right hand excites the string in an impulsive manner while the left hand keeps sustaining the tone as much as the construction of the instrument allows. In Task 2, participants were asked to alternate between single and double pitches in different takes. Finally, Task 3 presents a hybrid of the impulsive and sustained types. All given tasks focused on the notes B3 and C4 on the D string, played by index and middle fingers.

Each task was recorded as a fixed-form track, 2 min 16 s in duration, along with a metronome click at 70 BPM. The participants were instructed to play for 4 bars, rest for 2 bars, play the variation for 4 bars, rest another 2 bars and repeat this same 12-bar pattern two more times. See Table 1 for a detailed list of finger and style variations. To help the participants perform the tasks correctly, they were standing in front of a custom-built prompter screen. On the screen, they could follow animated circles, which signified the beat and the bar they were supposed to be at with respect to the predefined form of the given task. This allowed for a more comfortable and efficient experiment process. For the pilot study, we used a text-based prompting. However, this increased the cognitive load of the participants. Thus, for the full experiment we implemented a simple geometry-based design.

Table 1. Detailed Fingerings and Playing Styles Instructed to Participants for Particular Tasks.

	Takes 1-3-5	Takes 2-4-6
Impulsive	Index	Middle
Iterative	Index	Index–middle
Bending	Middle, as fast as possible	Middle, as slow as possible
Legato	Index–middle, hammer-on	Middle–index, pull-off

Note. Fingering and playing styles were organized based on the odd- and even-numbered takes to have a systematic approach to labeling different action features recorded within a single track. This approach facilitated the groupings of segmented individual takes during the preprocessing step.

Data Acquisition

Figure 3 shows the recording setup, which was based on two separate personal computers running the data collection software. In the first one, we used an external trigger to send the start pulse to the Qualisys motion capture system, which allowed an in-sync recording of the motion capture cameras, the Delsys Trigno EMG sensors, and the Canon video camera. The second computer recorded signals from the Myo armbands and the audio as line input from the guitar amplifier. This was accomplished using a custom-built Python program to record synchronized sensor data and audio. The Myo armbands were interfaced through improving the myo-to-osc framework for the Bluetooth API (Martin, Jensenius, & Torresen, 2018). To overcome possible bandwidth limitations, we implemented low-latency support for the multiple Myo armbands connected to the computer via individual Bluetooth Low Energy adapters. PyAudio was used for the audio recording (Pham, 2006). The Python interface ran as four simultaneous processes: data acquisition from each armband, the metronome, and the audio recording.

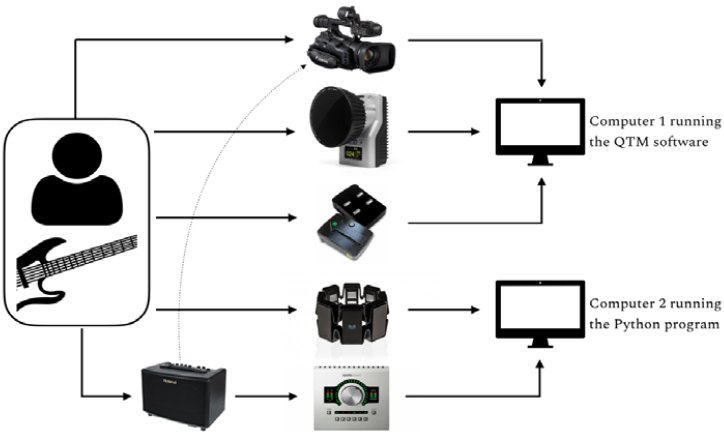


Figure 3. A simplified signal flow diagram of the experimental setup. Representative pictures of the equipment used, from top to bottom: Canon video camera, Qualisys Oqus infrared camera, Delsys Trigno electrodes, Myo armband, and Roland guitar amplifier, and Universal Audio Apollo Twin sound card.

Preprocessing

Preprocessing of our data for further analysis and modeling purposes was handled separately for the data from the Delsys and Myo systems. The medical-grade Delsys system provided high-quality data suitable for analytical purposes, while the Myo is a consumer-grade product that works well for interactive applications (see Pizzolato et al., 2017, for a comparison of various EMG acquisition setups). For the Delsys data, preprocessing included filtering, segmentation, and feature extraction methods. For the Myo data, we worked on interpolation and alignment of the raw data instead.

Synchronization

We synchronized the recorded data and audio through a custom-built metronome script within our Python program. This script recorded the timestamps of the metronome clicks together with the start point of the audio recording in a CSV file. This strategy helped in two ways. First, we could calculate lags at less than 0.1s among the various recording channels. As a result, we could align all the data types, based on their start points, to the metronome timeline. The synchronization strategy also helped in conforming the Qualysis data captured on Computer 1 with the line-audio recordings on Computer 2. Computer 1 ran the Qualisys software, which also recorded a standard video file synchronized with embedded audio.

We first extracted the audio stream from the video recording, and then decomposed the signal into its percussive and harmonic components. Applying an onset detection algorithm on the percussive component made it possible to obtain a timeline of metronome clicks from the ambient audio recording. This allowed us to measure the clicks and compare them to the logged timestamps of the original metronome clicks from Computer 2. Because the Delsys data shared the same timestamps with those of the metronome onsets, and the line audio recording shared the same timestamps with those of the metronome logs, we were able to align all the recorded data and media.

EMG Signal

Drawing on the method proposed by De Luca, Gilmore, Kuznetsov, & Roy (2010), we recorded the raw EMG data at 2000 Hz using the Delsys Trigno system, which were first run through a high-pass filter with a cutoff frequency of 20 Hz, and a low-pass filter with a cut-off of 200 Hz. Both filters were fourth-order Butterworth type (Selesnick & Burrus, 1998). Next, we segmented the synchronized and normalized EMG data into 5-beat sequences (1 bar created from the last beat of the previous bar in the timeline). This was to capture also muscle activation preceding the sound-producing action. The muscle activation necessarily precedes the motion of the hand and the audio onset.

Each task was recorded as a single track that contained six takes (see Table 1). Then, we selected one segment from each of them following this protocol:

1. Takes that featured the index finger on B3 were chosen from the impulsive and iterative tasks. In addition to an effort for narrowing the scope by focusing on the index finger for the impulsive task, we were interested in exploring how two motion types combine in the iterative task.

2. Takes that were played "as slow as possible" were chosen from the bending task. Slow bending (over a period of approximately a bar) is fairly similar to the sustained motion type. The guitar does not actually afford sustained performance in the same way as, for example, a violin does. However, the more the bending is prolonged, the more the damping is shortened. This results in two almost opposing input and output amplitude envelopes. The sustaining muscle amplitude envelope has an increased tension. The sound energy, on the contrary, decays quicker than that of an impulsive attack.
3. Takes that featured the hammer-on technique were chosen from the legato task. We observed that a majority of the participants was more comfortable with the hammer-on technique than a pull-off. This was also something we observed in the recorded data. In addition, hammer-on can be seen as a variation of the impulsive tasks played with both fingers.

Finally, each segment was divided into four EMG channels (i.e., the extensor and flexor muscles of each forearm). This resulted in 992 segments (31 participants, 8 tasks, 4 channels) of EMG data. Each segment had a duration of 4.29 s.

For the feature extraction, we were interested primarily in the amplitude envelopes. This was extracted as the root mean square (RMS) of the continuous signal. The moving RMS of a discrete signal is defined by St-Amant, Rancourt, & Clancy (1996) as

$$\hat{x}_1(t) = \left[\frac{1}{N}\sum_{i=t-N+1}^{t} m^2(i)\right]^{1/2}$$

where $\hat{x}$ is the EMG amplitude estimate at sample t, using a smoothing window length of N. The recommended window length for calculating the RMS of an EMG signal is 120–300 ms (Burden, Lewis, & Willcox, 2014). After several trials, we noticed that shorter window lengths better covered the peaks of fast attacks. Thus, we used a 50 ms sliding window with 12.5 ms (25%) overlaps.

Muscle onsets were calculated using the Teager-Kaiser Energy (TKE) operation to improve the accuracy of the detection (Li, Zhou, & Aruin, 2007). The TKE operation is defined in the time domain as

$$y(n) = x^2(n) - x(n-1)x(n+1)$$

Audio Signal

The sound analysis was based primarily on the RMS envelopes. Additionally, we computed the spectral centroid (SC) of the sound, as it has been shown to correlate with the perception of brightness in sound (Schubert, Wolfe, & Tarnopolsky, 2004), that is, how the spectral content is distributed between high and low frequencies. The RMS signal is particularly relevant in that our primary interest in this study is in the amplitude envelope of the sound. RMS correlates with perceptual loudness; people can judge whether a signal is loud, soft, or in between but cannot infer where a periodic signal is peaking or is at a zero-crossing (Beranek & Mellow, 2012; Ward, 1971). Thus, for our purposes, RMS served as an appropriate feature, providing more information than simply identifying the peak value within a given time interval.

Analysis

Our analysis focused on exploring similarities between the amplitude envelopes of the EMG signals and the sound. We achieved this by comparing the beginning and the end of the body–sound interactions identified when playing the electric guitar. Muscle activation was observable at the beginning, followed by motion, and then the resulting sound. We conducted the entire analysis through in a custom-built toolbox programmed in Python.

EMG Analysis

The initial component of the EMG analysis focused on exploring the similarities between the RMS of each of the four channels (two per arm) and the sound RMS for each of the participants. We used a Pearson's product–moment correlation, Spearman's rank correlation, and analysis of variance.

Also known as linear correlation coefficient (LCC), Pearson's product–moment correlation is a parametric correlation of the degree to which the change in one variable is linearly associated with a change in another continuous variable. In its equation form, LCC is commonly abbreviated as r while, in our case, x and y represent EMG and audio signals, respectively,

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{\sqrt{\sum(x - \overline{x})^2 \sum(y - \overline{y})^2}}$$

where $LCC > 0$ denotes a positive correlation while the opposite ($LCC < 0$) refers to an inverse correlation. The LCC approaches 0 when the correlation weakens. To our knowledge, this measure has not been used to compare audio and EMG signals.

A common assumption of the Pearson's correlation is that the continuous variables follow a bivariate normal distribution. In other cases, where the data is not normally distributed and the relationship of two variables rather seems nonlinear, the Spearman's rank correlation (SCC) is suggested to measure the monotonic relationship (Schober, Boer, & Schwarte, 2018). SCC is fairly similar to LCC, but it calculates the ranks of the pair of values. It is abbreviated as r_s (or ρ) in its mathematical representation where D is the difference between ranks and n denotes the number of data pairs:

$$r_s = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

A positive r_s denotes a covariance toward the same direction, whereas a negative r_s refers to fully opposite directions. It is a correlation measure that is commonly used in validating EMG data (Fuentes del Toro et al., 2019; Nojima, Watanabe, Saito, Tanabe, & Kanazawa, 2018).

A third approach was to calculate the pairwise *t* tests and one-way analysis of variance (ANOVA) to explore the variances of correlation values across participants and different dynamics. Here, we tested the assumptions of normality and homogeneity of variances of the independent samples in the dataset using the Shapiro-Wilk and Levene tests (Virtanen et al., 2020), respectively.

In addition to the above-mentioned analysis strategies, we explored other representations of the EMG signals. Inspired by Santello, Flanders, & Soechting (2002) and González Sánchez, Dahl, Hatfield, & Godøy (2019), we applied the time-varying Principal Component Analysis

(PCA) to merge all four channels and investigate prominent features across all participants. The input matrix for the PCA is defined as $A \in \mathbb{R}^{m \times n}$ where m is the number of participants and n denotes the number of EMG channels. For each of the 8 tasks, in which half employed soft dynamics and the other half strong dynamics, we obtained two principal components (PCs), which represented a combination of both excitation and modulation actions on the guitar, as shown by the following equation,

$$EMG_{\mathrm{m}} = \mathrm{meanEMG}_{\mathrm{m}} + PC1 \times EMG1_{\mathrm{m}} + \cdots + PCn \times EMGn_{\mathrm{m}}$$

Additionally, we applied Singular Spectrum Analysis (SSA) to principal components of EMG for further signal–noise separation. SSA is a technique of time series analysis used for decomposing the original series by means of a sliding window into a sum of small number of interpretable components, such as slowly varying trend, oscillatory (periodic) components, and structureless noise (Golyandina & Zhigljavsky, 2013). The algorithm for SSA is similar to that of PCA in multivariate data. In contrast to the PCA, which is applied to a matrix, SSA provides a representation of the given time series in terms of a matrix made of the time series (Alexandrov, 2009). In this way, we applied SSA on the EMG principal components and extracted the trend, which is a smooth additive component that contains information about the time series' global change (Alexandrov, Bianconcini, Dagum, Maass, & McElroy, 2012). This procedure allowed us to obtain better visualizations of the nonlinearity of relationships between EMG and audio waveforms.

It should be noted that researchers in the literature have suggested a variety of specialized methods for choosing the SSA window length (L). Knowing that it is highly difficult to define a universal method to find an optimal L value for an arbitrary time series and that the practitioners should therefore investigate this issue with care, Khan & Poskitt (2011) suggested a rule as $L = (\log N)^c$ with $c \in (1.5, 3.0)$ for assigning a window length that will yield near optimal performance. Starting from there, as the RMS segments of our interest were at a fixed length of $N = 344$, we empirically chose $c = 2.5$, which yielded $L = 10$.

Video Analysis

We used the Musical Gestures Toolbox (Jensenius, 2018b) to extract the sparse optical flow from the video recordings, with the goal of identifying to what extent participants moved unintentionally. This information allowed us to make comparisons with other data at hand and open a better understanding of unexpected muscle activations.

Sound Analysis

Our aim in the sound analysis was to quantify how the different dynamics influenced the overall brightness of the sound. To this end, we averaged the SC across all participants. Note that the sound data in this study is presented in approximately 4.29 s chunks. However, we also investigated chunks of a shorter duration in order to explore whether dynamic fluctuations of particularly the iterative task had an effect on the mean brightness. Moreover, considering the damping character of the guitar, which is relatively short in duration, we explored how decay times influenced the overall brightness value.

Results

The 36 participants completed 360 tasks in total. However, we excluded five datasets due to incomplete data. After also excluding the improvisations—which were intended to be used in the modeling experiment detailed below—we analyzed 248 tasks from 31 participants. An overview of how muscle activation patterns transform to sound features in each task is illustrated in Figure 4.

LCC and SCC

The correlation coefficients among participants were computed using the LCC and SCC measures. Table 2 shows positive correlation, negative correlation, mean, and standard deviation for each factor. Figures 5 and 6 show the distribution of LCC and SCC correlations.

The analysis shows to what extent the muscle activation underlying the sound-producing motion and the resultant sound on the same musical instrument can have similar amplitude envelopes. This is supported by the ANOVA results. The correlation of muscle–sound amplitude envelopes—whether positive, negative, or close to 0—does not exhibit a noteworthy variance between participants. That is, the ANOVAs for EMG–sound similarities across participants (for all EMG channels and tasks) are as follows: LCC, $F(30,961) = 1.6$, $p = 0.02$, and SCC, $F(30,961) = 1.59$, $p = 0.02$.

The comparisons of the correlation values between left and right hands supports the functional distinction between the right and left actions (see Table 3). Another clear distinction was revealed when we compared to what extent the EMG and sound envelopes correlated with respect to soft and strong dynamics (see Table 4). When the participants played strongly, the muscle and resultant sound amplitude envelopes correlated better.

PCA and SSA

Figure 7 shows the waveforms of the two principal components of the combined EMG channels across all participants for impulsive, iterative, bending, and legato tasks, separately for soft and strong dynamics. Each panel shows the activation patterns for the characteristics of these tasks.

The trends of the same principal component waveforms via signal–noise separation were extracted using SSA ($L = 10$) and have been plotted against the averaged sound RMS on the horizontal axis in Figure 8. Here we can observe the varying level of nonlinearities of the muscle–sound relationship for the tasks played at different dynamic levels.

Spectral Centroid

Figure 9 shows the distribution of the SC of the sound across all participants for each soft and strong task, separately. Although stronger dynamics show a clear strength in the upper end of the sound spectrum, the distribution among particular tasks varied depending on the chosen timescale. As such, SC values of all tasks with soft dynamics ($M = 299.03$, $SD = 124.24$), compared to the SC values of tasks with strong dynamics ($M = 585.93$, $SD = 141.22$), demonstrated significantly lower mass of the spectrum, $t(246) = 16.98$, $p < .001$

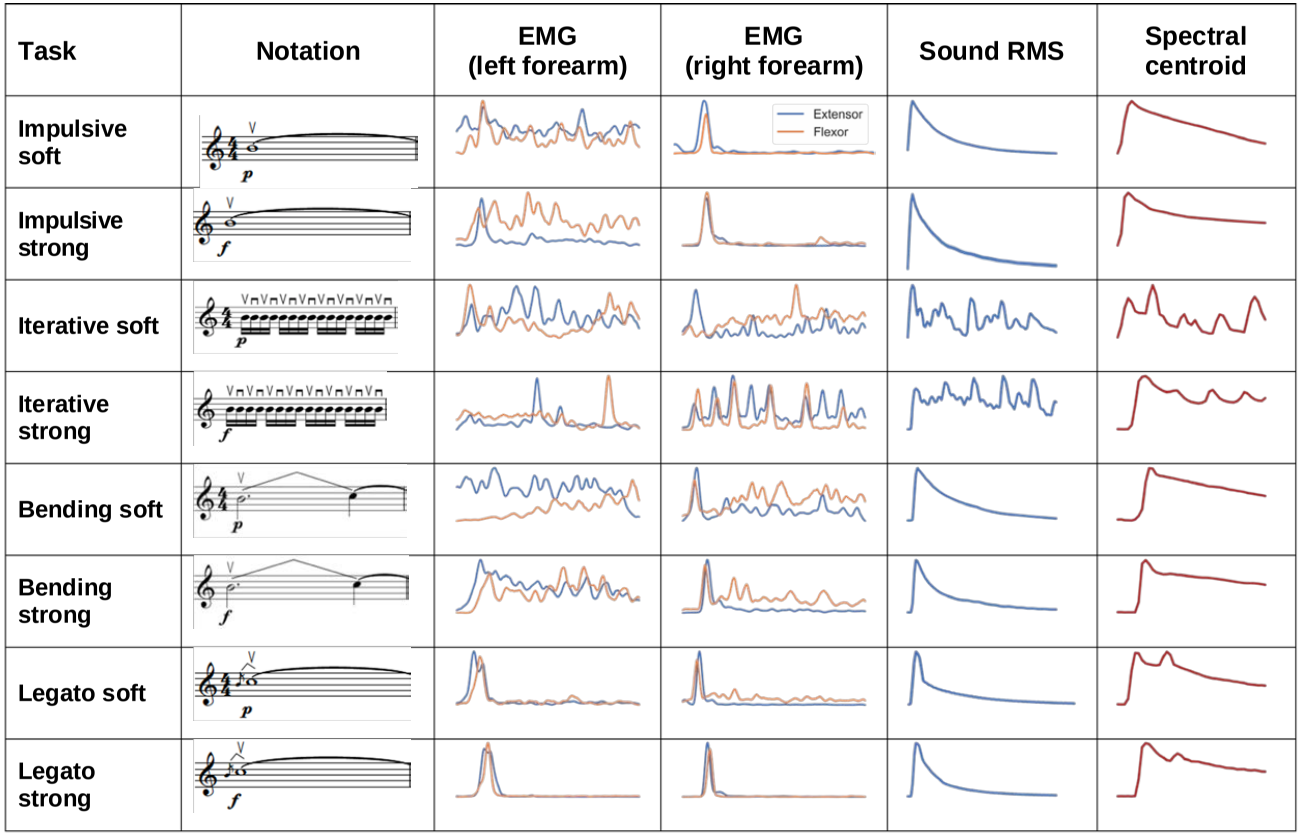


Figure 4. An overview of how notated music transforms into an audio waveform when playing the electric guitar. Trends of signals were extracted using Singular Spectrum Analysis (SSA) with a window length $L = 10$.

Table 2. Correlation Coefficients for Each Factor (LCC and SCC): The Positive, Negative, Mean and Standard Deviation of Correlation Coefficients.

			Impulsive soft	Impulsive strong	Iterative soft	Iterative strong	Bending soft	Bending strong	Legato soft	Legato strong
LCC	r	Extensor (right)	0.66	0.59	0.64	0.68	0.60	0.73	0.46	0.53
		Flexor (right)	0.65	0.54	0.51	0.86	0.65	0.69	0.42	0.55
		Extensor (left)	0.72	0.62	0.74	0.64	0.63	0.76	0.44	0.60
		Flexor (left)	0.55	0.55	0.65	0.65	0.48	0.63	0.51	0.48
	$-r$	Extensor (right)	–0.24	–0.03	–0.24	–0.24	–0.12	–0.10	–0.38	–0.24
		Flexor (right)	–0.34	–0.25	–0.10	–0.07	–0.34	–0.10	–0.33	–0.32
		Extensor (left)	–0.66	–0.61	–0.35	–0.35	–0.51	–0.66	–0.35	–0.33
		Flexor (left)	–0.62	–0.62	–0.53	–0.51	–0.54	–0.46	–0.30	–0.53
	μ	Extensor (right)	0.17	0.24	0.28	0.33	0.26	0.28	0.00	0.09
		Flexor (right)	0.13	0.23	0.22	0.33	0.21	0.27	0.02	0.03
		Extensor (left)	–0.23	–0.08	0.21	0.25	0.18	0.22	–0.02	0.01
		Flexor (left)	–0.34	–0.24	0.20	0.21	0.03	0.15	–0.01	–0.02
	σ	Extensor (right)	0.23	0.14	0.17	0.18	0.18	0.19	0.15	0.20
		Flexor (right)	0.25	0.17	0.17	0.19	0.21	0.17	0.13	0.18
		Extensor (left)	0.35	0.36	0.26	0.23	0.27	0.24	0.16	0.16
		Flexor (left)	0.28	0.25	0.28	0.20	0.14	0.22	0.14	0.12

(continued)

Table 2. Correlation Coefficients for Each Factor (LCC and SCC): The Positive, Negative, Mean and Standard Deviation of Correlation Coefficients. (continued)

			Impulsive soft	Impulsive strong	Iterative soft	Iterative strong	Bending soft	Bending strong	Legato soft	Legato strong
SCC	r_s	Extensor (right)	0.66	0.71	0.68	0.71	0.58	0.78	0.55	0.61
		Flexor (right)	0.49	0.71	0.58	0.74	0.66	0.74	0.27	0.66
		Extensor (left)	0.65	0.84	0.77	0.81	0.81	0.84	0.66	0.42
		Flexor (left)	0.70	0.70	0.69	0.63	0.43	0.70	0.43	0.34
	$-r_s$	Extensor (right)	–0.45	–0.15	–0.25	–0.30	–0.14	–0.17	–0.42	–0.33
		Flexor (right)	–0.41	–0.43	–0.18	–0.04	–0.41	–0.19	–0.19	–0.42
		Extensor (left)	–0.85	–0.89	–0.56	–0.56	–0.61	–0.85	–0.32	–0.61
		Flexor (left)	–0.77	–0.78	–0.50	–0.50	–0.62	–0.78	–0.55	–0.61
	μ	Extensor (right)	0.08	0.27	0.25	0.41	0.27	0.35	–0.01	0.10
		Flexor (right)	0.07	0.26	0.17	0.38	0.18	0.37	0.01	0.02
		Extensor (left)	–0.27	–0.08	0.27	0.35	0.19	0.25	0.00	0.00
		Flexor (left)	–0.38	–0.26	0.21	0.29	0.04	0.17	0.00	0.00
	σ	Extensor (right)	0.22	0.19	0.20	0.23	0.15	0.25	0.14	0.25
		Flexor (right)	0.24	0.21	0.19	0.19	0.18	0.25	0.12	0.20
		Extensor (left)	0.40	0.46	0.31	0.23	0.30	0.24	0.14	0.14
		Flexor (left)	0.31	0.31	0.31	0.23	0.16	0.26	0.13	0.10

Note. The zeros in the table represent rounded values that were smaller than three decimal places, thus a “close-to-zero” correlation.

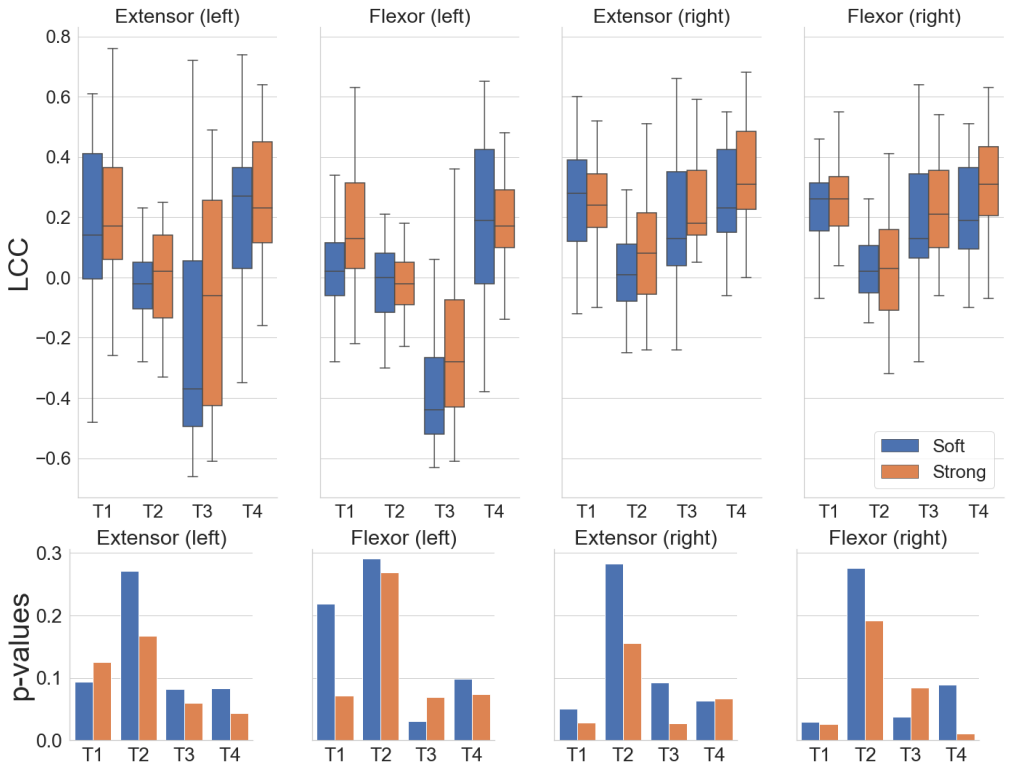


Figure 5. Pearson's product–moment correlations between EMG and Sound RMS envelopes. LCC > 0 denotes a positive correlation while LCC < 0 refers to the negative. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong dynamics. The bar plots below show the distribution of *p*-values showing the significance of the correlations. T1, T2, T3 and T4 refer to impulsive, iterative, bending and legato tasks, respectively.

Table 3. Pairwise *t* tests Demonstrating How Modification (Left Forearm) and Excitation (Right Forearm) Actions Have Distinct EMG–Sound Amplitude Envelopes.

	Modification action	**Excitation action**	**Variance**
LCC	$M = 0.03$, $SD = 0.30$	$M = 0.19$, $SD = 0.21$	$t(495) = 11.41$, $p < .001$
SCC	$M = 0.05$, $SD = 0.34$	$M = 0.20$, $SD = 0.24$	$t(495) = 9.04$, $p < .001$

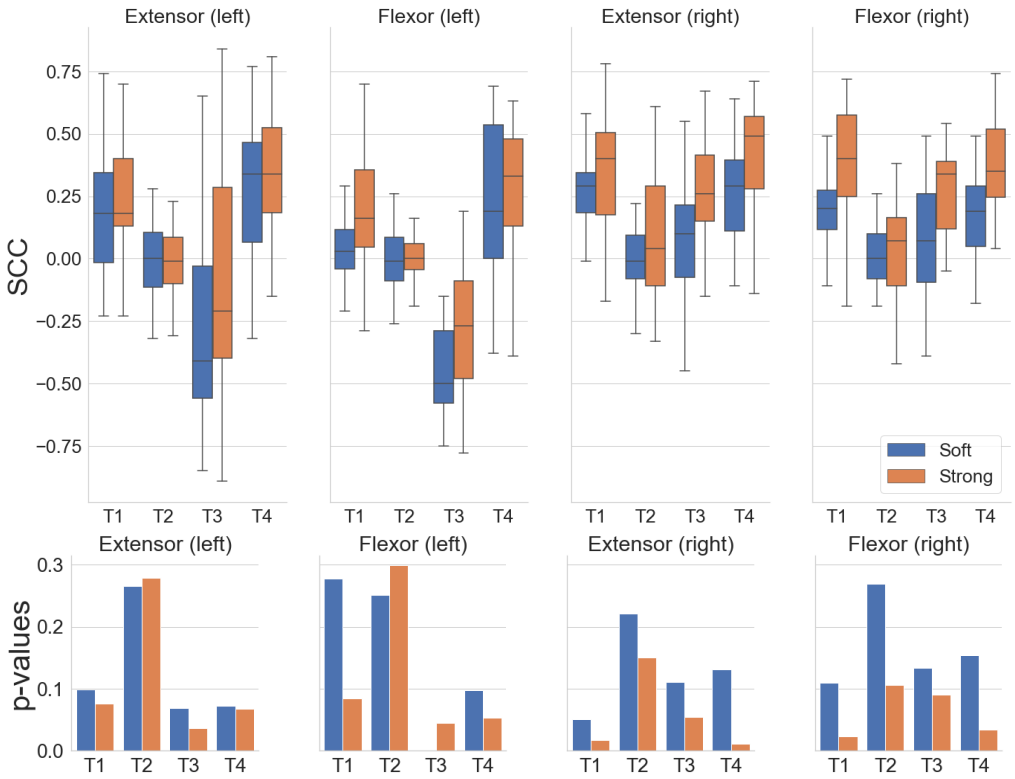


Figure 6. Spearman's rank correlations between EMG and Sound RMS amplitude envelopes. SCC > 0 denotes a covariance in the same direction while SCC < 0 refers to the opposite direction. The box plots show the interquartile ranges of correlation distribution per task, separately for soft and strong dynamics. The bar plots below show the distribution of *p*-values showing the significance of the correlations. T1, T2, T3 and T4 refer to impulsive, iterative, bending and legato tasks, respectively.

Table 4. Means, Standard Deviations and *t*-scores for LCC and SCC Metrics.

	Soft	Strong	Variance
LCC	$M = 0.08$, $SD = 0.27$	$M = 0.14$, $SD = 0.26$	$t(495) = 5.41$, $p < .001$
SCC	$M = 0.07$, $SD = 0.29$	$M = 0.18$, $SD = 0.31$	$t(495) = 8.33$, $p < .001$

Note. Pairwise t-tests show EMG–sound amplitude envelopes correlations between soft and strong dynamics.

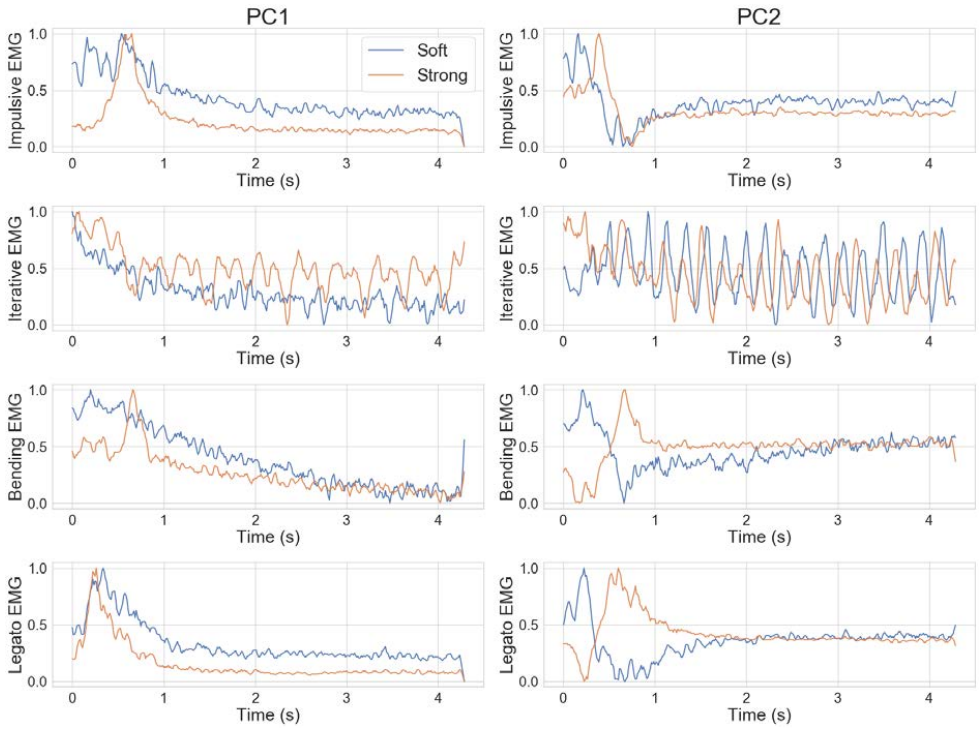


Figure 7. Two principal components (PC1 and PC2) of the combined left and right forearm EMG data of all participants rescaled to (0,…,1) (See the text for more information about the PCA analysis).

Discussion

The analyses showed that sound production on musical instruments is a phenomenon that involves many physical and physiological processes. For example, Figure 10 shows the activation patterns of the extensor and flexor muscles during down- and up-stroking using a plectrum. This figure illustrates only two muscles groups from the right forearm. However, a musical note often is produced as a more complex combination of both arms, as shown in Figure 4.

Similarity Between EMG and Sound Shapes

Our experiment results show that the relations between the muscle energy envelope and the envelope of the resultant sound have similarities between participants. The results show a significant variance when comparing attacks with soft and strong dynamics using pairwise *t*-tests (Table 4). As shown in Figures 5 and 6, the correlation values are higher, and the directionality is more apparent when the same task is played with strong dynamics. This may be due to two factors. First, greater energy input results in larger sound amplitude, which is less biased to base noises, such as the inherent postural instability of the human body.

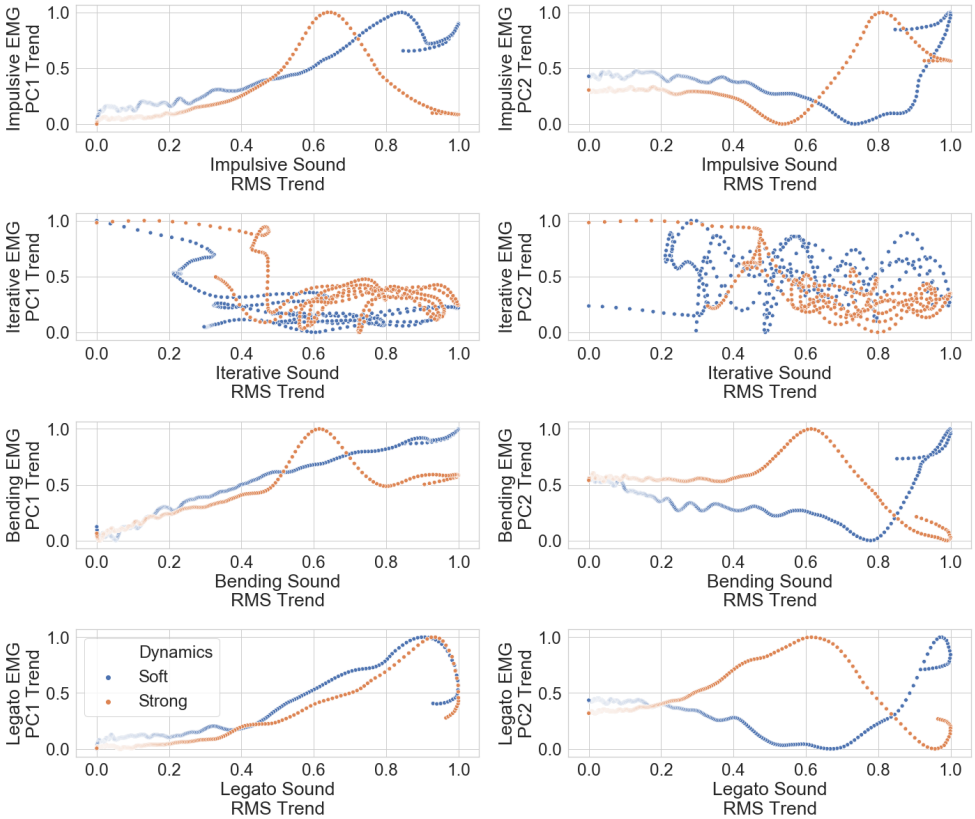


Figure 8. Decomposed principal components (PC1 and PC2) against resultant Sound RMS of all participants (SSA window length *L = 10*). The plots show to what extend the EMG and resultant sound RMS envelopes have a linear relationship at every time step.

Second, we know that expert players tend to use less tension in the forearm muscles (Winges, Furuya, Faber, & Flanders, 2013). Most of our participants can be considered semiprofessionals and thus may have felt less comfortable with stronger dynamics. As a result, they may have employed forearm muscles more explicitly. Unfortunately, we do not have data to check this hypothesis.

The results in Table 3 are in line with the conceptual distinction provided in our Introduction. The excitation action, which typically is performed by the right arm for right-handed players, determines the main characteristics of the resultant sound amplitude envelope. The difference between the activation patterns of both forearms is also observable in Figure 4. The impulsive tasks noted on the top two rows, for example, show the right forearm muscles have envelopes similar to that of the resultant sound while the activation patterns from the left forearm seem to resemble a continuous sound envelope, somewhat between the sustained and iterative types. This is due mainly to a continuous effort exerted by the left forearm over the period of the given task, which is different from the right forearm that excites the string once,

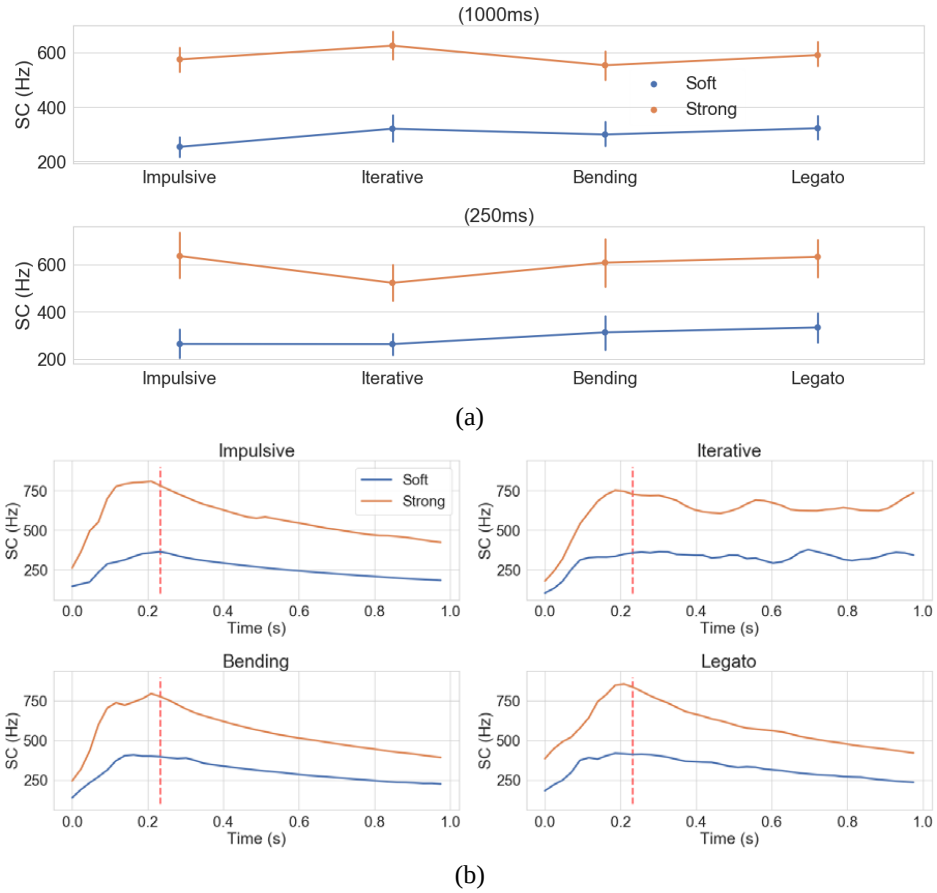


Figure 9. Spectral centroid (SC) of the resultant sound (a) SC distribution between soft and strong dynamics in chunks of 1000 ms and 250 ms duration. (b**)** SC envelopes averaged across all participants. The red vertical lines on the left sides of the plots show the cut point of 250 ms. Note that the segments are 1 s long, which is different than 4 s segments that we initially used. Doing so removed most of the decay that contributes to mean SC.

exerting effort for just a short period. During continuous exertion, we see that bioelectric muscle signals do not exhibit a smooth trend yielding a nearly iterative shape.

Furthermore, any additional ancillary motion, such as moving parts of the body to the beat, or a further modification motion, such as a vibrato to add expression to the sustaining tone, also can be considered as possible artifacts contributing to the envelope of muscular activation. When inspecting the individual participants' video recordings, we noticed that such spontaneous motions are fairly common. Figure 11 provides an example of this. We extracted the sparse optical flow by tracking certain points on a close-up video recording of a participant playing the impulsive task. The participant's ancillary motion is observable in the position of the guitar in relation to the camera and captured possibly by the EMG sensors on the left forearm.

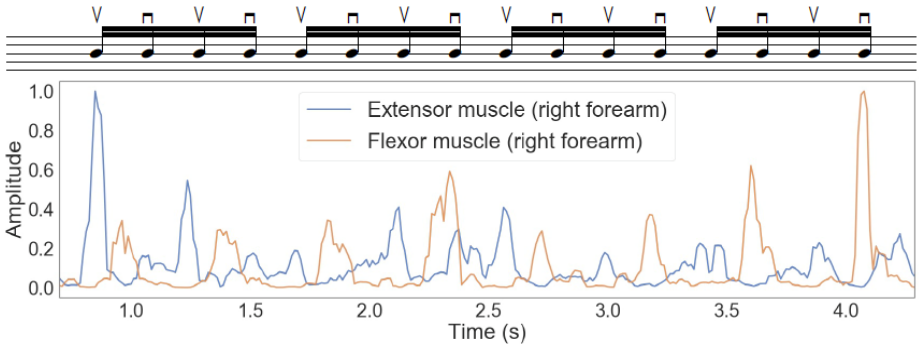


Figure 10. EMG amplitude of the excitation motion during iterative task demonstrating distinct activation of extensor and flexor muscles for down and up strokes, respectively, during a series of 16th notes.

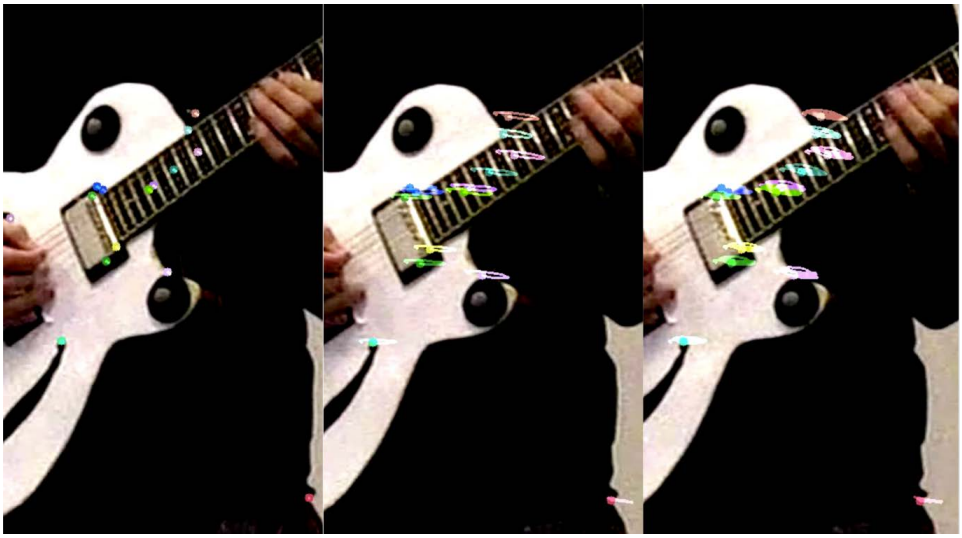


Figure 11. The sparse optical flow shows the trajectory of multiple points on a close-up video segment while a participant is performing an impulsive task. Three subsequent screenshots demonstrate the ancillary motion reflected on the guitar over the period of 1 bar (~3.43 s). The multicolored points on the left picture yield certain patterns in their trajectories reflecting participant movement patterns in the center and right pictures.

We suggest that such ancillary motion influences more directly the ongoing muscle activation as compared to right forearm muscles, which were resting at that moment.

When comparing left and right forearm muscle activation patterns, the negative directionality is noteworthy. This is particularly clear during the bending tasks (see Figures 5 and 6), a playing technique in which the right arm excitation is equivalent to the impulsive task. The left arm modifies the pitch and has a sustained envelope. This is unique to the guitar, as this instrument

does not afford sustained sound as do the bowed strings instruments. We should also mention that both the exerted effort and the resultant damping character of the sound would be different if other equipment were used, such as a harder wood and/or pickups with stronger magnets in instrument design, high-gain amplifiers, electronic effects units, or any other room acoustics resulting in greater feedback.

Another interesting observation when comparing data from the left and right forearms is the similarity between positive correlation values of the Impulsive and Legato. This could result from coarticulation. In this task, the left hand executes two consecutive (impulsive) attacks. These are quite different from the impulsive task, however. Because the two consecutive attacks are close temporally, they merge to form one large, coarticulated shape.

Finally, the iterative tasks showed the most idiosyncratic patterns and the least shape similarity. We observed that playing consecutive notes as a series of relatively fast attacks was the most challenging task for many of our participants. Depending on the level of expertise, each participant demonstrated signs of slogging to some extent, which arguably resulted in unique timing characteristics. Effort constraints may be a relevant topic here: Although some players are able to optimize their muscle contractions, others can exert more or less than optimal effort. In addition to the participants' level of expertise, the iterative task may have led to muscle fatigue. None of the participants mentioned this condition, but the possibility deserves further exploration in the context of musical performance.

Exploring Dimensions

The main objective of this investigation was to explore the quantifiable similarities of the amplitude envelopes of sound-producing actions on the electric guitar. In the first part of our analysis, we explored such relationships between two muscle groups against the resultant sound amplitude envelopes from each participant. In the second, we focused on a combination of results from all muscles on both forearms across all participants. We performed PCA on concatenated EMG channels, aiming to render additional observations and visual perspectives. In this part of the analysis, then, we aimed at exploring the signal PCs that can reflect a combination of simultaneous processes. Our interpretation of the PCA is that although PC1 reflected the overall dissipating aspect of the excitation motion, PC2 revealed the variation in the energy input of the modulation motion. This is the case even though we did not specify the decomposition to be separate.

From these observations, we can group all types of EMG patterns under two conceptual categories: (a) impulsive, where a single impulse or a series of impulses is applied, and (b) sustained, denoting a constant muscle energy. The experimental approach of decomposing the PCs using SSA (Figure 8) provided alternative perspectives for exploring the nonlinearities of the relationships. Whereas series of impulses yielded fewer regular patterns, sustaining energy showed clearer similarities. These findings are in line with the results presented in the previous subsection.

The Resultant Sound

Figure 9a demonstrates how SC was distributed across various tasks and dynamics. The main observation here was that stronger dynamics led to a brighter sound. We also should note that plucked strings have what may be called incidental nonlinearities that can have effects, depending on the intensity of excitation (Fletcher, 1999). Moreover, we used 1000 ms and 250 ms segments

in these two subplots, respectively. These durations were different from the approximately 4.29 s segments we relied on in our analysis. This shift was intended to remove the tail of the waveform during the decay, which affects the mean brightness value. So, our results support previous work suggesting that timescales shorter than 500 ms reflect most of the timbral features that happen during the attack phase of the excitation (Godøy, 2018).

Figure 9a shows how Iterative had a brighter character than the others when the averaged segments are a longer duration (1000 ms). However, Iterative's mean SC decreased when shorter segments (250 ms) were used for comparison. This indicated a timbral difference between the impulsive and iterative tasks. That is, the impulsive tasks tended to demonstrate a single peak in the exerted energy, reflecting in a brighter sound. The series of attacks of the latter, however, showed more fluctuating energy. This also revealed that during those series, the energy that was transduced into the attacks also made the SC change dynamically. As such, the plots of the averaged SC shaped over time (Figure 9b).

EXPERIMENT 2: A PRELIMINARY PREDICTIVE MODEL

Following the empirical exploration of how biomechanical energy transforms into sound, we used these transformations as part of a machine learning framework based on a long short-term memory recurrent neural network for action–sound mappings. We engaged an interdisciplinary approach that draws on a combination of sound theory and embodied music cognition. Our starting point involved an idea of developing a model that is trained solely on fundamental sound-producing action types. The aim this component of our research was to predict the sound amplitude envelopes of a freely improvised performance. We see this as a preliminary step toward designing an entirely new instrument concept.

Conceptual Design

Our motivating concept was to develop a model that allows for coadaptation, meaning the system not only learns from the user but the user adapts to the behavior of the system (Tanaka & Donnarumma, 2018). Knowing that EMG is a stochastic and nonstationary signal (Phinyomark, Campbell, & Scheme, 2019), even simple trigger actions are quite complex in nature. Although it may seem handy to use well-known machine learning methods, such as classification for triggering sounds or regression to map continuous motion signal (Caramiaux & Tanaka, 2013), we are interested in developing beyond a one-directional control. This vision is conceptually different from, for example, using machine learning for EMG-based control aimed at prosthetic research (Jaramillo-Yánez, Benalcázar, & Mena-Maldonado, 2020).

We also were intrigued with another design concept: predictive modeling. Following various control structures that we had explored in previous work (Erdem, Camci, & Forbes, 2017; Erdem & Jensenius, 2020; Erdem, Schia, & Jensenius, 2019), we were interested more with the ways of how the system can behave differently from interactive music systems that react primarily to the user (Rowe, 1992). Drawing on the work of Martin, Glette, Nygaard, & Torresen (2020), we began exploring the potential of artificial intelligence tools generally, and predictive models in particular, that facilitate not only the input–output mapping of complex signals in new instruments but also enable self-awareness.

Methods

Data Preparation

Our modeling process relied heavily on data from Myo armbands, as they are a cheaper and more portable solution than the Delsys Trigno system. As described in detail in the Methods section of Experiment 1, we synchronized the EMG data and audio arrays based on the recorded metronome timeline. The primary difference in our analysis procedure in this experiment was that we kept all data for modeling. That is, the data were not segmented nor did we eliminate the material collected in-between tasks, when the participants were waiting for the next instruction. This latter set of material made it possible to have the model learn to distinguish between rest and motion states.

We applied linear interpolation to the EMG data and calculated the RMS from the audio signal. The data preparation process resulted in eight segments per participant of EMG and audio data as training examples. The preliminary architecture focused on mapping the raw EMG data to the RMS envelope of the sound as the target.

Predictive Model

We used nine model configurations based on a long short-term memory (LSTM) recurrent neural network (RNN) architecture. Drawing on previous research that suggested 32 or 64 LSTM units in each layer as the most appropriate for integrating the model into an interactive music system (Martin & Torresen, 2019), we wanted to test different configurations. Thus, we used models with one, two, and five hidden layers and each containing 16, 32, and 64 units. Each model was trained on sequences that were 50 data points. This window size refers to 250 ms at Myo armband's 200 Hz sample rate.

Following the LSTM layer(s), a fully connected layer passes a single data point into the activation layer, using a rectified linear activation (ReLU) function. From there, a final layer returns the mean value of the input tensor in order to map an EMG window to one data point of the sound RMS, a many-to-one sequence modeling problem. In short, an array of raw EMG signal with a dimensionality of (50,16) was fed into the network as sliding windows (e.g., sample N_0 to N_{49}, sample N_1 to N_{50}, etc.) to predict a single value of sound RMS at a time step (see Figure 12 for a simplified diagram). The training loss function was defined as

$$\mathcal{L}(x_{\mathrm{RMS}}, \hat{x}_{\mathrm{RMS}}) \quad = \frac{1}{n}\sum_{i=1}^{n}(x_{\mathrm{RMS,i}} - \hat{x}_{\mathrm{RMS,i}})^2 ,$$

where x_{RMS} are the recorded values, $\hat{x}_{\mathrm{RMS}}$ are the values to be predicted, and the sliding window has size n.

Training

The dataset was limited to 160 training examples from 20 participants in which 40 examples were used for validation. We conducted the training using the Adam optimizer (Kingma & Ba, 2014) with a batch size of 100. As we executed multiple trainings to test various configurations, we limited the trainings to 20 epochs. The duration of trainings varied from 4 to 10 hours, depending

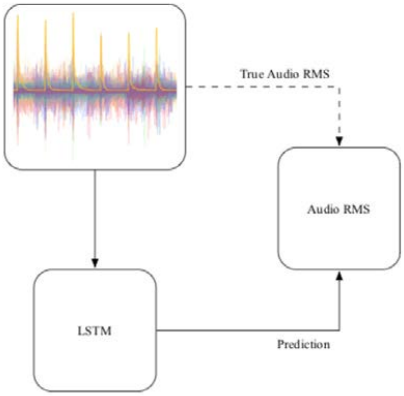


Figure 12. Sketch of the training model: A 16-channel Raw EMG as the source and sound RMS as the target data are passed into an LSTM cell, which then outputs a prediction.

on the quantity of trainable parameters in relation to the number of hidden layers and units. Even though we report here the final results from training locally on a single Nvidia GeForce GTX 1080Ti graphics processing unit (GPU), we also ran the trainings on Google's browser-based coding notebook, *Colaboratory;* we did not observe any remarkable difference in the training duration.

Results

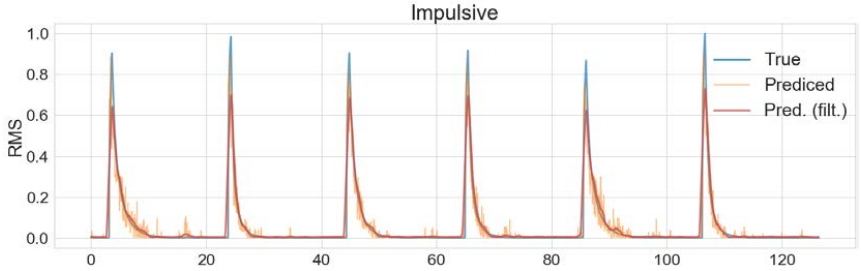
All model configurations were generally capable of predicting the sound RMS (see Figure 13). The model with two hidden layers and 64 units had the best results, which can be seen in the figures of recorded versus predicted RMS of the impulsive (Figure 13a) and iterative tasks (Figure 13b). For the latter, the model could generate similar consecutive envelopes resembling a series of attacks.

One goal in developing this preliminary model was to test the performance of the LSTM based on a limited dataset. In this case, the limitation refers to the type of dataset rather than its size. We were encouraged to see that the model could predict the general trend of the sound energy when tested using the free improvisation dataset (Figure 14).

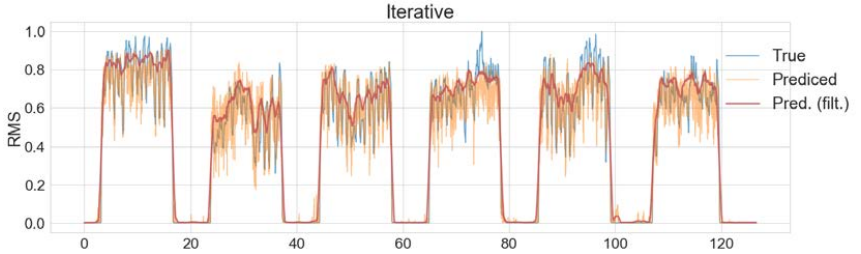
The prediction of the bending task brought an interesting result (Figure 13c). Normal guitar performance does not afford sustained excitation action, although it can be accomplished with a bow on the strings, as Led Zeppelin's guitarist, Jimmy Page, popularized in the late 1960s. However, apart from using extended playing techniques—such as pressing on the strings with the hands or using additional equipment, such as a bow, vibrato arm, or electronic effects processing units—a player can only hit on a string once (impulsive) or as a series of impulses (iterative). Thus, sustained motion is available only for the modification action, such as bending the string with a finger on the left hand.

In the prediction, however, we observed a longer decay as compared to an impulsive, single attack of the right arm. This interesting in-between result suggests a means for augmenting the guitar for creative purposes.

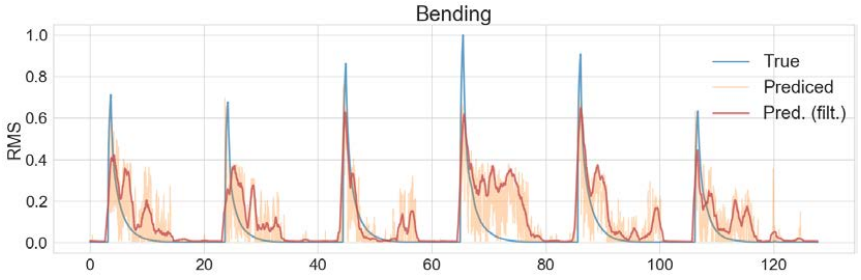
We also tested various model sizes using Euclidean distance measure (EDM), which is a common method for measuring the distance between objects. EDM is calculated as the root of square



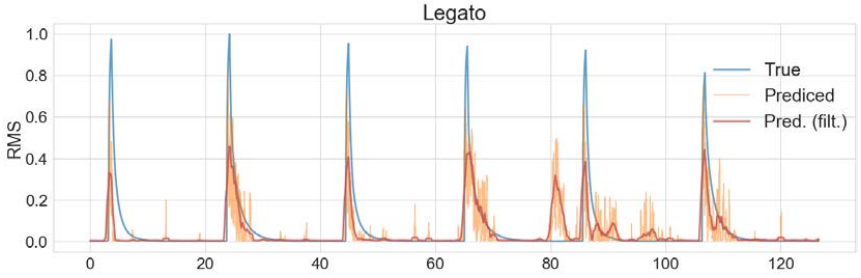
(a) The RMS of the recorded sound and the model prediction for the impulsive task.



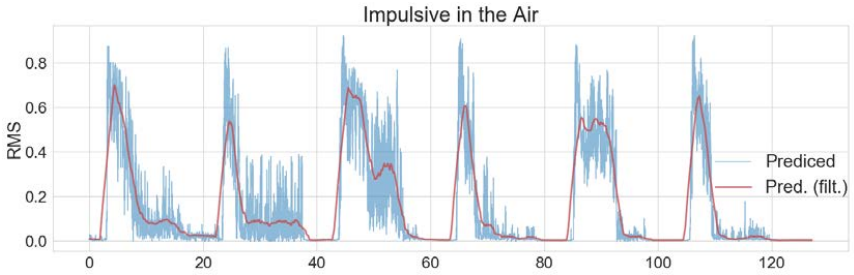
(b) The RMS of the recorded sound and the model prediction for the iterative task.



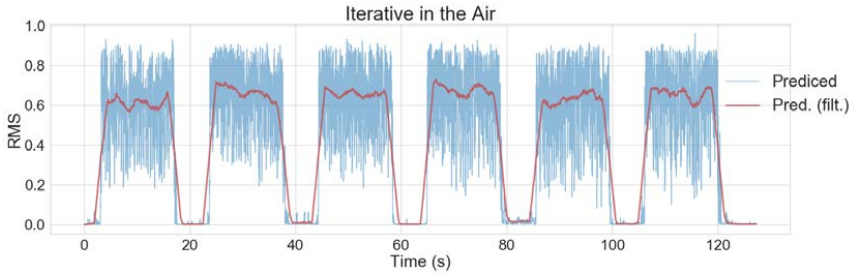
(c) RMS of the recorded sound and the model prediction for the bending task.



(d) RMS of the recorded sound and the model prediction for the legato task.



(e) The predicted sound RMS of impulsive playing in the air.



(f) The predicted sound RMS of iterative playing in the air.

Figure 13. The performance of the model with two hidden layers and 64 units in given tasks. Plots a through d show the true sound RMS and predicted RMS envelopes. Because we recorded impulsive and iterative tasks performed in the air as test data for further exploration, plots e and f show only the predicted sound RMS envelope based on the EMG data of an air performance. The time axis is shared across all plots and predicted curves are processed with a Savitzky-Golay filter (Savitzky & Golay, 1964) to reflect the general shape and facilitate the visual inspection.

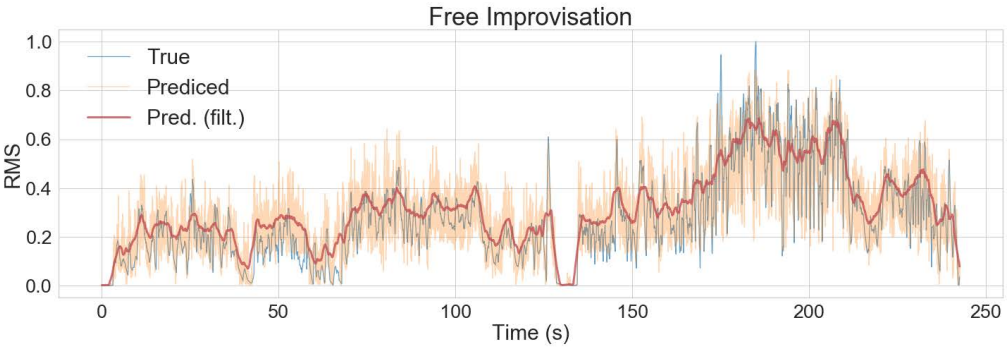


Figure 14. The RMS of the recorded sound and the model prediction of a free improvisation task. Predicted curves are filtered to reflect the general shape and facilitate the visual inspection.

differences between coordinates of two objects (Kang, Cheng, Lai, Shiu, & Kuo, 1996). Given the normalized true and predicted sound RMS vectors $\vec{p}, \vec{s} \in \mathbb{R}^n$, we can find the distances in Euclidean n-space as $\sqrt{(p_1 - s_1)^2 + (p_2 - s_2)^2 \ldots (p_n - s_n)^2}$. The distances between the true RMS and predicted RMS envelopes of the nine models of different configurations were calculated using the free improvisation recordings from 20 participants, of which given tasks were used as training data. This provided us with a statistical measure for evaluating the performance of different model configurations for mapping 16-channel raw EMG data to sound RMS envelope. Figure 15 provides the distribution of distances together with the latency of single-threaded prediction processes on the central processing unit (CPU) of a MacBook Pro 2018. According to results, we observed a trend that the model performance increases along with additional LSTM layers and units; unfortunately, however, the model's performance decreases when the model becomes too large. The prediction time also increases drastically with additional parameters. However, models with a single hidden layer have the least latency even while having a fairly large margin of error. Thus, according to the results, a two-layer stacked LSTM with 32 or 64 units can be seen as a "sweet spot" configuration.

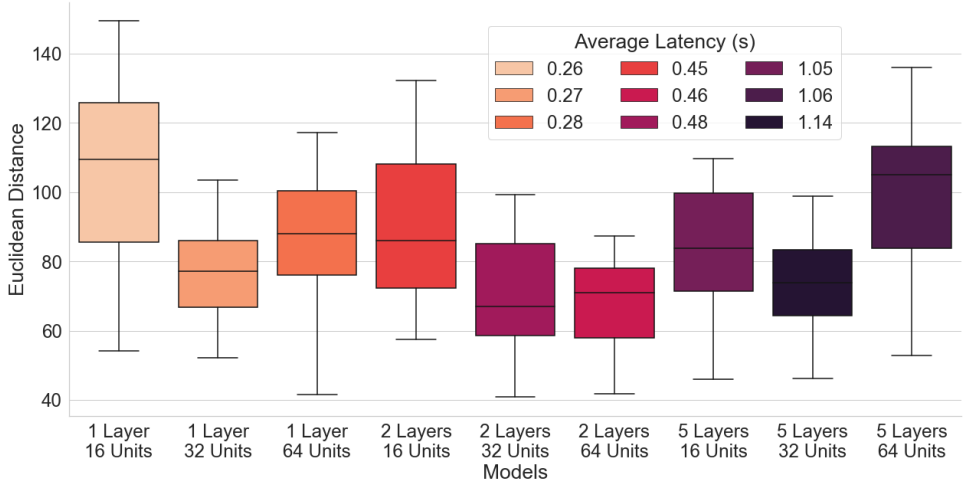


Figure 15. Euclidean distances between true RMS envelope of the free improvisation task and its corresponding prediction of RMS envelope based on nine model configurations. The boxes display the interquartile ranges while the central lines show the median. The whiskers show the minimum and maximum values of the distribution.

Discussion

The implemented model can predict the overall trend of the sound energy of a freely improvised performance based solely on a training dataset of particular action types. As shown in Figure 13, some similarities are evident between the EMG signal and the sawtooth-like patterns of the predicted waveforms. We think this is acceptable, as these fluctuating patterns can be filtered easily and used as an amplitude parameter in the sound synthesis. However, considering that

the prediction of a single temporal feature is insufficient for capturing the complexity of musical sound, these patterns might cause problems. These predictions also may lead to unpredictable sound features that could be aesthetically pleasing in an improved model.

Drawing on the results from the tests between different model configurations, we see that, as the model size increases, the distance between the true RMS and predicted RMS generally decreases, but the similarity tends to increase. However, larger model sizes also result in a larger latency, which can cause problems in real-time performance situations. We believe that although a lower similarity can be utilized creatively, higher similarity with a larger latency is much less usable.

Another step in the future development of the system will be to conduct a thorough user study to test the framework. It will be particularly interesting to explore how possible it is to obtain near-optimal latency using the trained model and, moreover, how to use the latency creatively. Also relevant is the exploration of how motion data from an inertial measurement unit can add to the information provided by the EMG data. At its core, the question remains how the spatiotemporality of the performance can be further explored and evaluated.

GENERAL DISCUSSION AND CONCLUSIONS

The main research question that inspired the first experiment of the study regarded the relationships between action and sound in instrumental performance. To answer that, we performed statistical analyses on the data from an experiment in which 31 electric guitarists performed a set of basic sound-producing actions: impulsive, sustained, and iterative. The results showed clear action–sound correspondences, compatible with theories of embodied music cognition. These correspondences' statistical levels varied, depending on the given task. The relatively less-challenging tasks, such as impulsive, yielded higher correlation values. Conversely, we observed how participants' varying level of motor control resulted in unique EMG and audio wave-forms for the iterative tasks, which involved performing a series of impulsive sound-producing actions merged into a single shape. Here, the way participants used rhythms and structured the musical time had a determinant role in the coarticulated muscle activations. Thus, we can argue that complex rhythms yield unique bodily patterns.

An important limitation of Experiment 1 was the gender imbalance. Unfortunately, only one female joined the study. The participants were recruited via local communication channels; thus the range of participants was limited to whoever volunteered. Another limitation was the experimental setup in a controlled laboratory environment, which may have felt unnatural to many participants. The same could be said about the very constrained tasks, which restricted the participants' musical expression. For example, the use of physical effort is most likely quite different than in a live music-making situation. Also, we provided the participants with the instrument, which may have influenced the results. Musicians typically develop bodily habits based on particular instruments—including the string gauge and plectrum. Thus, unfamiliarity with the electric guitar used in this study could have affected the relationships between EMG and audio signals. Furthermore, the analyses clearly showed that these relationships contain nonlinear components, so we could question the reliability of using linear methods. Still, we believe that the use of such methods can provide an example for future work. The results were satisfactory

for such an exploratory study, but the choice of statistical methods for correlating bodily signals with sound features remains an open question.

The second research question involved how such relationships between action and sound can be used to create new instrumental paradigms. Relying on the notion of imitating existing interactions in new instruments, we aimed in our second experiment at modeling the action–sound relationships found in playing the guitar. We explored some aspects of this question through a series of analyses in the first experiment. However, we were more focused in Experiment 2, employing our multimodal dataset to train LSTM networks of different configurations. Our results showed that the preliminary models could predict audio energy features of free improvisations on the guitar, relying on an EMG dataset of three distinct motion types. These results satisfied our expectations concerning the size and type of the training dataset. Considering the nonlinear components found in the analysis of the relationships between the EMG and sound RMS envelopes (see Figure 8), the satisfactory outcome of our model corresponded to the known ability of neural networks that, in theory, any continuous function can be approximated by computing the gradient through a neural network. This is achieved by breaking down a complex function into several step-functions computed by the network's hidden neurons. How good the approximation is often depends on the depth or number of layers in the network and the width or number of neurons of each layer (Goodfellow et al., 2016).

A caveat of our research in our second experimental setup is that even the smallest model configuration achieved a much higher latency (see Figure 15 for the results of our analysis on different model configurations) than acceptable ranges (20–30 ms) for real-time audio applications (Lago & Kon, 2004). Although it is possible to reduce the latency using elaborated programming structures, a single predicted feature would still be limited. Moreover, a similar output can be achieved using traditional signal processing methods. Thus, a next step in our research will include expanding the model with spectral, temporal, and spatial features from both motion and audio data. It would also be relevant to explore the potential of what such a deep learning-based framework can afford for musical performance and creativity in a new instrumental concept.

In the future, we will continue to build on this two-fold strategy of combining empirical data collection and machine learning-based modeling. We intend to explore deep learning features for myoelectric control that can be applied to extracting discriminative representations of coarticulated sound-producing actions. We remain interested especially in exploring the creative potential of such models: How can artificial intelligence generally—and deep neural networks particularly—be used to explore the aesthetics of, and embodied interaction with, the transformations of biomechanical waveforms into sound? To answer such a question, we will emphasize exploring the conceptual and practical challenges of space and time, particularly when using the human body as part of the musical instrument. By conducting more user studies, we expect to provide valuable information about conceptual approaches of translating embodied knowledge of actions into the use of new musical instruments.

IMPLICATIONS FOR RESEARCH

The studies presented in this paper are situated within the interdisciplinary research field of music technology (see Serra, 2005). This field involves both practitioners and researchers working with both artistic and scientific methods. Both groups will benefit from the knowledge gained from our

empirical studies of basic sound-producing actions and the artificial intelligence methods developed for modeling relationships between muscle energy and audio energy. More broadly, the outcomes of applying multimodal machine learning for creative purposes opens new research activities. These contributions include a new multimodal dataset, the development of custom software tools, statistical analyses between action and sound, and an evaluation of various machine learning configurations. Furthermore, the study provides additional support for previous research on action–sound relationships and embodied music cognition. Our emphasis on EMG irregularities as a control signal suggests an alternative perspective for music technology research on performing arts and human-computer interaction. These irregularities and imperfections open for new creative possibilities.

REFERENCES

Alexandrov, T. (2009). *A method of trend extraction using singular spectrum analysis.* Retrieved from https://arxiv.org/abs/0804.3367

Alexandrov, T., Bianconcini, S., Dagum, E. B., Maass, P., & McElroy, T. S. (2012). A review of some modern approaches to the problem of trend extraction. *Econometric Reviews, 31*(6), 593–624. https://doi.org/10.1080/07474938.2011.608032

Beranek, L. L., & Mellow, T. J. (2012). Chapter 1: Introduction and terminology. In L. L. Beranek & T. J. Mellow (Eds.), *Acoustics: Sound fields and transducers* (p. 1–19). Cambridge, MA, USA: Academic Press. https://doi.org/10.1016/B978-0-12-391421-7.00001-4

Briot, J.-P., Hadjeres, G., & Pachet, F.-D. (2020). *Deep learning techniques for music generation*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-70163-9

Burden, A. M., Lewis, S. E., & Willcox, E. (2014). The effect of manipulating root mean square window length and overlap on reliability, inter-individual variability, statistical significance and clinical relevance of electromyograms. *Manual Therapy, 19*(6), 595–601. https://doi.org/10.1016/j.math.2014.06.003

Cadoz, C., & Wanderley, M. M. (2000). Gesture-music. In M. M. Wanderly & M. Battier (Eds.), *Trends in gestural control of music* (Vol. 12, pp. 71–94). Paris, France: IRCAM. Retrieved from https://hal.archives-ouvertes.fr/hal-01105543

Caramiaux, B., Bevilacqua, F., Zamborlin, B., & Schnell, N. (2009). Mimicking sound with gesture as interaction paradigm (Technical Report). Paris, France: IRCAM. Retrieved from http://articles.ircam.fr/textes/Caramiaux10d/index.pdf

Caramiaux, B., & Donnarumma, M. (2020). *Artificial intelligence in music and performance: A subjective art-research inquiry*. Retrieved from https://arxiv.org/abs/2007.15843

Caramiaux, B., Montecchio, N., Tanaka, A., & Bevilacqua, F. (2015). Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems, 4*(4), 18–52. https://doi.org/10.1145/2643204

Caramiaux, B., & Tanaka, A. (2013). Machine learning of musical gestures: Principles and review. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 513–518). Daejeon, Republic of Korea: Zenodo. https://doi.org/10.5281/zenodo.1178490

De Luca, C. J., Gilmore, L. D., Kuznetsov, M., & Roy, S. H. (2010). Filtering the surface EMG signal: Movement artifact and baseline noise contamination. *Journal of Biomechanics, 43*(8), 1573–1579. https://doi.org/10.1016/j.jbiomech.2010.01.027

Donnarumma, M. (2015). Ominous: Playfulness and emergence in a performance for biophysical music. *Body, Space & Technology, 14*, unpaginated. http://doi.org/10.16995/bst.30

Dreyfus, H. L. (2001). Phenomenological description versus rational reconstruction. *Revue Internationale de Philosophie, 216*(2), 181–196. https://doi.org/10.3917/rip.216.0181

Erdem, C., Camci, A., & Forbes, A. (2017). Biostomp: A biocontrol system for embodied performance using mechanomyography. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 65–70). Copenhagen, Denmark: Zenodo. http://doi.org/10.5281/zenodo.1176175

Erdem, C., & Jensenius, A. R. (2020). RAW: Exploring control structures for muscle-based interaction in collective improvisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 477–482). Birmingham, UK: Birmingham City University.

Erdem, C., Schia, K. H., & Jensenius, A. R. (2019). Vrengt: A shared body–machine instrument for music–dance performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 186–191). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672918

Fiebrink, R. A. (2011). *Real-time human interaction with supervised learning algorithms for music composition and performance* (Doctoral dissertation, Princeton University). Retrieved from https://www.cs.princeton.edu/research/techreps/TR-891-10

Fiebrink, R. A., & Caramiaux, B. (2016). *The machine learning algorithm as creative musical tool*. Retrieved from https://arxiv.org/abs/1611.00379

Fletcher, N. H. (1999). The nonlinear physics of musical instruments. *Reports on Progress in Physics, 62*(5), 723–764. http://doi.org/10.1088/0034-4885/62/5/202

Françoise, J. (2015). *Motion-sound mapping by demonstration* (Doctoral dissertation, Université Pierre et Marrie Curie). Retrieved from https://www.julesfrancoise.com/documents/JulesFRANCOISE_phdthesis.pdf

Fuentes del Toro, S., Wei, Y., Olmeda, E., Ren, L., Guowu, W., & Díaz, V. (2019). Validation of a low-cost electromyography (EMG) system via a commercial and accurate EMG device: Pilot study. *Sensors, 19*(23), 1–16. https://doi.org/10.3390/s19235214

Fyans, A. C., & Gurevich, M. (2011). Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 495–498). Oslo, Norway: Zenodo. http://doi.org/10.5281/zenodo.1178019

Godøy, R. I. (2006). Gestural-sonorous objects: Embodied extensions of Schaeffer's conceptual apparatus. *Organised Sound, 11*(2), 149–157. https://doi.org/10.1017/S1355771806001439

Godøy, R. I. (2017). Key-postures, trajectories and sonic shapes. In D. Leech-Wilkinson & H. M. Prior (Eds.), *Music and Shape* (pp. 4–29). New York, NY, USA: Oxford University Press. https://doi.org/10.1093/oso/9780199351411.003.0002

Godøy, R. I. (2018). Sonic object cognition. In R. Bader (Eds.), *Springer handbook of systematic musicology* (pp. 761–777). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-55004-5

Godøy, R. I., Haga, E., & Jensenius, A. R. (2005). Playing "air instruments": Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, & J. F. Kamp (Eds.), *International gesture workshop: Gesture in human–computer interaction* (pp. 256–267). Berlin, Germany: Springer. http://dx.doi.org/10.1007/11678816_29

Golyandina, N., & Zhigljavsky, A. (2013). *Singular spectrum analysis for time series*. Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-34913-3

González Sánchez, V. E., Dahl, S., Hatfield, J. L., & Godøy, R. I. (2019). Characterizing movement fluency in musical performance: Toward a generic measure for technology enhanced learning. *Frontiers in Psychology, 10*, 1–11. https://dx.doi.org/10.3389/fpsyg.2019.00084

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA, USA: The MIT Press. Retrieved from https://www.deeplearningbook.org/

Google. (2020). *Teachable machine*. https://teachablemachine.withgoogle.com/

Gritten, A., & King, E. (Eds.). (2011). *New perspectives on music and gesture*. London, UK: Routledge. https://doi.org/10.4324/9781315598048

Hatten, R. S. (2006). A theory of musical gesture and its application to Beethoven and Schubert. In A. Gritten & E. King (Eds.), *Music and gesture* (pp. 1–23). London, UK: Routledge. https://doi.org/10.4324/9781315091006

Hunt, A., & Wanderley, M. M. (2002). Mapping performer parameters to synthesis engines. *Organised Sound*, *7*(2), 97–108. https://doi.org/10.1017/S1355771802002030

Ingold, T. (2000). *The perception of the environment*. London, UK: Routledge.¨ https://doi.org/10.4324/9780203466025

Jaramillo-Yánez, A., Benalcázar, M. E., & Mena-Maldonado, E. (2020). Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. *Sensors*, *20*(9), Art. 2467. https://doi.org/10.3390/s20092467

Jensenius, A. R. (2007). *ACTION–sound: Developing methods and tools to study music-related body movement* (Doctoral dissertation, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-18922

Jensenius, A. R. (2017). Sonic microinteraction in "the air." In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 431–439). New York, NY, USA: Routledge. https://doi.org/10.4324/9781315621364

Jensenius, A. R. (2018a). Methods for studying music-related body motion. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 805–818). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-55004-5

Jensenius, A. R. (2018b). *The musical gestures toolbox for matlab*. In the *Late-Breaking/Demo Session Abstracts for the 2018 International Society for Music Information Retrieval Conference*. Retrieved from http://www.arj.no/wp-content/2018/09/Jensenius-ISMIR2018.pdf

Jensenius, A. R., & Lyons, M. J. (2017). *A nime reader: Fifteen years of new interfaces for musical expression*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-47214-0

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. 12–35). New York, NY, USA: Routledge. https://doi.org/10.4324/9780203863411

Kamen, G. (2013). Electromyographic kinesiology. In D. G. E. Robertson, G. E. Caldwell, J. Hamill, G. Kamen, & S. N. Whittlesey (Eds.), *Research methods in biomechanics* (2nd ed., pp. 179–202). North Yorkshire, UK: Human Kinetics, Inc.

Kang, W.-J., Cheng, C.-K., Lai, J.-S., Shiu, J.-R., & Kuo, T.-S. (1996). A comparative analysis of various EMG pattern recognition methods. *Medical Engineering & Physics*, *18*(5), 390–395. https://doi.org/10.1016/1350-4533(95)00065-8

Karjalainen, M., Mäki-Patola, T., Kanerva, A., & Huovilainen, A. (2006). Virtual air guitar. *Journal of the Audio Engineering Society*, *54*, 964–980. Retrieved from http://users.spa.aalto.fi/mak/PUB/AES12860.pdf

Kelkar, T. (2019). *Computational analysis of melodic contour and body movement* (Doctoral dissertation, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-74166

Khan, M. A. R., & Poskitt, D. S. (2011, November). *Window length selection and signal-noise separation and reconstruction in singular spectrum analysis* (Working Paper 23/11). Retrieved from the Monash University Department of Econometrics and Business Statistics website: http://business.monash.edu/econometrics-and-business-statistics/research/publications/ebs/wp23-11.pdf

Kiefer, C. (2014). Musical instrument mapping design with echo state networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 293–298). London, UK: Zenodo. http://doi.org/10.5281/zenodo.1178829

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. Retrieved from https://arxiv.org/pdf/1412.6980.pdf

Knapp, R. B., & Lusted, H. S. (1990). A bioelectric controller for computer music applications. *Computer Music Journal*, *14*(1), 42–47. https://doi.org/10.2307/3680115

Kozak M., Nymoen K., & Godøy R. I. (2012). Effects of spectral features of sound on gesture type and timing. In E. Efthimiou, G. Kouroupetroglou, & S. E. Fotinea (Eds.), *Gesture and sign language in human–computer interaction and embodied communication* (pp. 69–80). *Lecture Notes in Computer Science*, Vol, 7206. Berlin, Germany: Springer. o

Lago, N. P., & Kon, F. (2004). The quest for low latency. In *Proceedings of the International Computer Music Conference* (pp. 33–36). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.1143

Lee, M., Freed, A., & Wessel, D. (1991). Real-time neural network processing of gestural and acoustic signals. In *Proceedings of the International Computer Music Conference* (pp. 277–281). Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1991.064

Leman, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA, USA: The MIT Press. https://doi.org/10.7551/mitpress/7476.001.0001

Li, X., Zhou, P., & Aruin, A. S. (2007). Teager–kaiser energy operation of surface emg improves muscle activity onset detection. *Annals of Biomedical Engineering*, *35*(9), 1532–1538. https://doi.org/10.1007/s10439-007-9320-z

Maes, P.-J., Leman, M., Lesaffre, M., Demey, M., & Moelants, D. (2010). From expressive gesture to sound. *Journal on Multimodal User Interfaces*, *3*(1-2), 67–78. https://doi.org/10.1007/s12193-009-0027-3

Magnusson, T. (2019). *Sonic writing: Technologies of material, symbolic, and signal inscriptions*. London, UK: Bloomsbury Academic. https://doi.org/10.1080/14794713.2020.1765577

Martin, C. P., Glette, K., Nygaard, T. F., & Torresen, J. (2020). Understanding musical predictions with an embodied interface for musical machine learning. *Frontiers in Artificial Intelligence*, *3*, Art. 6. https://doi.org/10.3389/frai.2020.00006

Martin, C. P., Jensenius, A. R., & Torresen, J. (2018). Composing an ensemble standstill work for myo and bela. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 196–197). Blacksburg, VA, USA: Zenodo. http://doi.org/10.5281/zenodo.1302543

Martin, C. P., & Torresen, J. (2019). An interactive musical prediction system with mixture density recurrent neural networks. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 260–265). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672952

Mendoza Garay, J. I., & Thompson, M. (2017). Gestural agency in human-machine musical interaction. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 431–439). New York, NY, USA: Routledge. https://doi.org/10.4324/9781315621364

Nojima, I., Watanabe, T., Saito, K., Tanabe, S., & Kanazawa, H. (2018). Modulation of EMG-EMG coherence in a choice stepping task. *Frontiers in Human Neuroscience*, *12*, Art. 50. https://doi.org/10.3389/fnhum.2018.00050

Nymoen, K., Caramiaux, B., Kozak, M., & Torresen, J. (2011). Analyzing sound tracings: A multimodal approach to music information retrieval. In *Proceedings of the International ACM Workshop on Music Information Retrieval with User-centered and Multimodal Strategies*, (pp. 39–44). New York, NY, USA: ACM. https://doi.org/10.1145/2072529.2072541

Næss, T. R. (2019). *A physical intelligent instrument using recurrent neural networks* (Master's thesis, University of Oslo). Retrieved from http://urn.nb.no/URN:NBN:no-73901

Paine, G. (2009). Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, *14*(2), 142–155. https://doi.org/10.1017/S1355771809000259

Pakarinen, J., Puputti, T., & Välimäki, V. (2008). Virtual slide guitar. *Computer Music Journal*, *32*(3), 42–54. https://doi.org/10.1162/comj.2008.32.3.42

Pham, H. (2006). *Pyaudio: Portaudio v19 python bindings*. Retrieved from https://people.csail.mit.edu/hubert/pyaudio

Phinyomark, A., Campbell, E., & Scheme, E. (2019). Surface electromyography (EMG) signal processing, classification, and practical considerations. In G. Naik (Ed.), *Biomedical signal processing* (pp. 3–29). Singapore: Springer. https://doi.org/10.1007/978-981-13-9097-5_1

Pizzolato, S., Tagliapietra, L., Cognolato, M., Reggiani, M., Müller, H., & Atzori, M. (2017). Comparison of six electromyography acquisition setups on hand movement classification tasks. *PloS One*, *12*(10), e0186132. https://doi.org/10.1371/journal.pone.0186132

Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, *13*(2), 206–219. https://doi.org/10.1109/JSTSP.2019.2908700

Rowe, R. (1992). *Interactive music systems: Machine listening and composing*. Cambridge, MA, USA: The MIT Press. Retrieved from https://wp.nyu.edu/robert_rowe/text/interactive-music-systems-1993/

Santello, M., Flanders, M., & Soechting, J. F. (2002). Patterns of hand motion during grasping and the influence of sensory guidance. *Journal of Neuroscience*, *22*(4), 1426–1435. https://doi.org/10.1523/JNEUROSCI.22-04-01426.2002

Schacher, J. C., Miyama, C., & Bisig, D. (2015). Gestural electronic music using machine learning as generative device. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 347–350). Baton Rouge, Louisiana, USA: Zenodo. http://doi.org/10.5281/zenodo.1179172

Schaeffer, P. (2017). *Treatise on musical objects* (C. North & J. Dack, Trans.). Oakland, CA, USA: University of California Press. https://doi.org/10.1525/9780520967465-001 (Original work published in 1967)

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proceedings of the International Conference on Music Perception and Cognition* (pp. 654–657). Retrieved from http://newt.phys.unsw.edu.au/~jw/reprints/SchWolTarICMPC8.pdf

Selesnick, I. W., & Burrus, C. S. (1998). Generalized digital Butterworth filter design. In *IEEE Transactions on Signal Processing*, *46*(6), 1688–1694. https://doi.org/10.1109/78.678493

Serra, X. (2005). Towards a roadmap for the research in music technology. In *Proceedings of the International Computer Music Conference.* Retrieved from https://repositori.upf.edu/handle/10230/34486?locale-attribute=en

Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, *2*(2), 107–126. https://doi.org/10.1017/S1355771897009059

St-Amant, Y., Rancourt, D., & Clancy, E. A. (1996). Effect of smoothing window length on rms emg amplitude estimates. In *Proceedings of the IEEE Annual Northeast Bioengineering Conference* (pp. 93–94). New Brunswick, NJ, USA: IEEE. https://doi.org/10.1109/NEBC.1996.503233

Tahiroğlu, K., Kastemaa, M., & Koli, O. (2020). Al-terity: Non-rigid musical instrument with artificial intelligence applied to real-time audio synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 331–336). Birmingham, UK: Birmingham City University.

Tanaka, A. (1993). Musical technical issues in using interactive instrument technology with application. In *Proceedings of the International Computer Music Conference* (pp. 124–126). Tokyo, Japan: ICMC. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1993.023

Tanaka, A. (2015a). Intention, effort, and restraint: The EMG in musical performance. *Leonardo Music Journal*, *48*(3), 298–299. https://doi.org/10.1162/LEON_a_01018

Tanaka, A. (2015b). *Myogram* [Music composition and performance]. Retrieved from https://youtu.be/G6H1J2k--5I

Tanaka, A. (2019). Embodied musical interaction. In S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, & M. Wanderley (Eds.), *New directions in music and human−computer interaction* (pp. 135–154). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-92069-6

Tanaka, A., Donato, B. D., Zbyszynski, M., & Roks, G. (2019). Designing gestures for continuous sonic interaction. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 180–185). Porto Alegre, Brazil: Zenodo. http://doi.org/10.5281/zenodo.3672916

Tanaka, A., & Donnarumma, M. (2018). The body as musical instrument. In Y. Kim & S. L. Gilman (Eds.), *The Oxford handbook of music and the body*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190636234.013.2

Valles, M. L., Martínez, I. C., Ordás, M. A., & Pissinis, J. F. (2018). Correspondence between the body modality of music students during the listening to a melodic fragment and its subsequent sung interpretation [Abstract]. In *Proceedings of th 15th International Conference on Music Perception and Cognition & the*

10th Triennial Conference of the European Society for the Cognitive Sciences of Music (p. 308). Retrieved from http://sedici.unlp.edu.ar/handle/10915/70462

Van Nort, D., Wanderley, M. M., & Depalle, P. (2014). Mapping control structures for sound synthesis: Functional and topological perspectives. *Computer Music Journal*, *38*(3), 6–22. https://doi.org/10.1162/COMJ_a_00253

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J. Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Visi, F., Coorevits, E., Schramm, R., & Miranda, E. R. (2017). Musical instruments, body movement, space, and motion data: Music as an emergent multimodal choreography. *Human Technology*, *13*(1), 58–81. https://doi.org/10.17011/ht/urn.201705272518

Waisvisz, M. (1985). The hands, a set of remote midi-controllers. In *Proceedings of the International Computer Music Conference* (pp. 313–319). Burnaby, BC, Canada: ICMC. Retrieved from http://hdl.handle.net/2027/spo.bbp2372.1985.049

Ward, M. R. (1971). *Electrical engineering science*. New York, NY, USA: McGraw-Hill.

Winges, S. A., Furuya, S., Faber, N. J., & Flanders, M. (2013). Patterns of muscle activity for digital coarticulation. *Journal of Neurophysiology*, *110*(1), 230–242. https://doi.org/10.1152/jn.00973.2012

Authors' Note

The authors thank the participating musicians, as well as Victor Evaristo González Sánchez and Julian Führer, for their contributions during the data collection and modeling processes. This work was supported in part by the Research Council of Norway (Project 262762) and NordForsk (Project 86892).

All correspondence should be addressed to
Çağrı Erdem
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion
Department of Musicology
University of Oslo
Postboks 1133 Blindern 0318 Oslo, Norway
cagri.erdem@imv.uio.no

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi

Paper V

Tool or Actor? An Evaluation of a Musical AI "Toddler" with Two Expert Improvisers

Çağrı Erdem, Benedikte Wallace, Kyrre Glette, Alexander Refsum Jensenius

Submitted in *Computer Music Journal*.
September 2021

V

Tool or Actor? An Evaluation of a Musical AI "Toddler" with Two Expert Improvisers

Firstname Lastname

Anonymous.

Abstract

In this paper we introduce the coadaptive audiovisual instrument CAVI. This instrument uses deep learning to generate its control signals based on muscle and motion data of the performer's actions. The generated control signals automate the live sound processing based on layered time-based effects modules. How is such an instrument perceived by the performer? Is it an instrument or an actor? We report on an evaluation of CAVI and its use in a public event with two expert improvisers. The evaluation is based on interviews with the performers and questionnaires filled out by audience members. The analyses showed that whether such an instrument is experienced as a tool or actor is closely linked with the performer's sense of agency, which varies throughout a performance depending on several factors, such as perceived qualities of the musical coordination, a delicate balance between surprising and familiar elements, and physical aspects of the performance environment.

«BEGIN ARTICLE»

Imagine playing your electric guitar while someone else is tweaking the knobs of the effects pedals. According to studies investigating the sense of agency, such situations create ambiguity in one's sensed control over her actions. New interfaces for musical expression (NIMEs) have employed a variety of machine learning (ML) techniques for

action–sound mappings since the early 1990s (Lee et al. 1991). Over the last decades, there has also been more interest in researching musical *agents* within the broader field of artificial intelligence (AI) and music (Collins 2006). Agent comes from the Latin word *agere*, meaning "to do" (Russell 2010). Essentially any person or thing that acts purposefully might be considered an agent. For example, an agent's sole purpose can be to recognize repeating pitch intervals (Minsky 1981). Such an artificial agent is concerned with tackling a musical task, hence be a *musical agent*. Traditionally, instruments (except for the human voice) have been physical objects with sound-producing mechanical properties. New music technologies allow various types of musical agency. However, what does it take for an instrument to "act" like an agent in music interaction? This is a question that has been discussed by several authors, including Launay (2015); Mendoza and Thompson (2017); Dahlstedt (2021). However, there are few studies that investigate expert musicians' experiences with technological musical agents.

This article is focused on an evaluation of the coadaptive audiovisual instrument CAVI, which generates its own control signals based on the performer's previously executed actions. CAVI builds on a laboratory study of guitarists' sound-producing actions and a mapping model developed from the empirical data (Erdem et al. 2020). The idea has been to develop a system that lets an acoustic performer improvise with automated live sound processing. CAVI has a generative model that continuously receives muscle activation (EMG) and acceleration (ACC) signals from the performer and predicts a new set of such data akin to the most likely action the performer would take. The live sound processing uses time-based effects, exploring a complexity spectrum between temporal ambiguity and familiar actions. The model obscures causality by blending acoustic and electro-acoustic sounds. We have been interested in answering the following questions:

- Can musical coordination emerge between a human acoustic performer and a

generative machine-based instrument?

- How do expert musicians experience such a system with respect to control and agency?

We start by presenting some information regarding the background of the project, its artistic origins, and some key concepts. Then follows an explanation of the system architecture and the evaluation performed with both performers and audience members after a concert.

Background

CAVI builds on a dataset collected in a previous study of the sound-producing actions of guitarists (Erdem et al. 2020). This dataset consists of muscle activation (EMG) and accelerometer (ACC) data and audio recordings of thirty-three guitarists playing a number of basic sound-producing actions (impulsive, sustained, and iterative) and free improvisations. The long short-term memory (LSTM) recurrent neural network (RNN) model we developed was satisfying in capacity; it could predict the sound energy envelope of improvised recordings based on a training dataset of solely basic actions. However, the predictions were subject to perceivable latency and a weakened sense of agency.

Sense of Agency

Several studies have stressed negative impact of temporal incongruities on the experience of agency (Haggard et al. 2002; Ebert and Wegner 2010; Kawabe 2013). The sense of agency is defined as one's sense of control over the consequences of actions (Moore 2016). Latency is one example of the loss of sense of agency that many electroacoustic musicians have experienced. Another example is the acoustic feedback loops that may occur when using microphones in front of speakers. Such feedback is

unwanted in many cases, but can also be used creatively. Some musicians use feedback actively in experimental electroacoustic music (Liontiris 2018; Melbye and Ulfarsson 2020). According to Kiefer et al. (2020), such a feedback instrument has "a life of its own" by means of circulating signals. Sensorimotor accounts of the sense of agency suggest an internal comparator mechanism called *forward prediction model* (Wolpert et al. 1995), which checks the congruency of the signal generated by the motor system when performing an action and incoming sensory signals (e.g., auditory, proprioceptive). The result yields the sense of agency or the lack thereof (Haggard 2005; Jeannerod 2008; Gentsch and Schutz-Bosbach 2015).

Emergent Coordination

Studies that account for perceptual cues have shown that agency judgments could also rely on perceptual influences in passive conditions (Knoblich and Repp 2009). They could also be based onthe perceived quality of a shared performance (van der Wel et al. 2012). Examples of instruments or experimental music practices that involve varying levels of the loss of control include various multi-user NIMEs (Weinberg and Gan 2001; Fels et al. 2004; Kaltenbrunner et al. 2006). One common aspect of these works is what Knoblich et al. (2011) call "emergent coordination," in which coordinated behavior of multiple agents arises without a plan. Free improvisation is a performance practice that is often based on such emergent coordination (Bailey 1993; Borgo 2005; Kosowitz and Vickery 2013) or a collaboratively emergent character (Sawyer and DeZutter 2009).

Human–Machine Improvisation

Diverse approaches have been taken for emerging coordination between humans and machines in improvisation settings. In the *coadaptive* control paradigm, it is not only the system that reacts to the user but the user is also expected to adapt to the system's behavior (Tanaka and Donnarumma 2018). One early example is the MIDI-controlled

agents of *Voyager* by Lewis (2000), which collect musical details played by the performer and reproduce them with both surprise and familiarity within a rule-based structure. According to Lewis, Voyager is an example of a "de-instrumentalized" computer system. Similarly, *FILTER* by Nort et al. (2013) aims to achieve an intelligence state of "careful" listening akin to free improvisation and sound-based electroacoustic aesthetics. *MASOM* by Tatar and Pasquier (2017) prioritizes similar aesthetics, using a cognitive model of "sound affect estimation" proposed by Russell (1980). This way, MASOM generates abstract, yet meaningful, sounds and noises regarding the performer's measured affective states. Finally, *AI-terity* by Tahiroglu et al. (2021) is a non-rigid NIME using a generative model for sound synthesis. This can be seen as an example of an instrument somewhere between a tool and an autonomous agent.

CAVI

The main goal of CAVI was to develop a coadaptive instrument. It was inspired by the works mentioned above but differs in several ways. From a sensing perspective, CAVI is inherently multimodal. It is based on muscle and motion data as well as sound. As opposed to our predictive model from a previous project (Erdem et al. 2020), CAVI is based on a generative framework that makes predictions by sampling from a probability distribution. An analogy would be that while our former system tries to guess the ingredients of a dish, the latter does its best to re-cook from the taste it learned from examples.

System Architecture

The generative modeling approach that CAVI takes is based on mixture density networks (MDNs). Such MDNs treat the outputs of a neural network as the parameters of a Gaussian mixture model (GMM) (Ellefsen et al. 2019), which, according to Martin and Torresen (2019), are suitable for modeling music improvisation processes. A GMM can be

derived using the mean, weight and standard deviation of each component. A mixture density recurrent neural network (MDRNN) can be formed when an MDN is combined with a recurrent neural network (RNN), with which we can make real-valued predictions based on a sequence of inputs. The system architecture of CAVI is based on such MDRNNs, which have been successful in projects such as speech recognition (Schuster 1999), handwriting (Graves 2013), and drawing sketches (Ha and Eck 2017).

Figure 1 depicts a simplified signal flow of CAVI's performance system including the MDRNN used in this work. The RNN consists of two layers of long short-term memory (LSTM) cells (Schmidhuber 2009). The LSTM layers contain 64 hidden units each. The outputs of the second LSTM layer are in turn connected to an MDN. The LSTM layers learn to estimate the mean (μ), standard deviation (σ) and weight (π) of the five Gaussian distributions of the MDN. The number of components needed to accurately represent the data is not known and is treated as a hyperparameter. In our case, the GMM consists of $K = 5$ n-variate Gaussian distributions. The model is trained using the Adam optimizer (Kingma and Ba 2014) until the loss on the validation set failed to improve for 20 consecutive epochs. This approach has the advantage of control over the diversity and "randomness" of sampling, and control over the number of mixture components that allow training to account for situations where multiple predictions could be considered equally suitable.

Approach

CAVI is inspired by the way Martin (2019) uses an MDRNN framework in call-and-response mode. The model "re-cooks" whatever it learns from the given data. For example, it can generate how you likely would carry on with the melody you started playing if trained on a dataset of the songs you usually play. That resembles the call-and-response systems developed in jazz contexts, such as in the *Continuator* of Pachet (2003). However, in Martin's framework, the model is trained on a motion dataset. Thus,

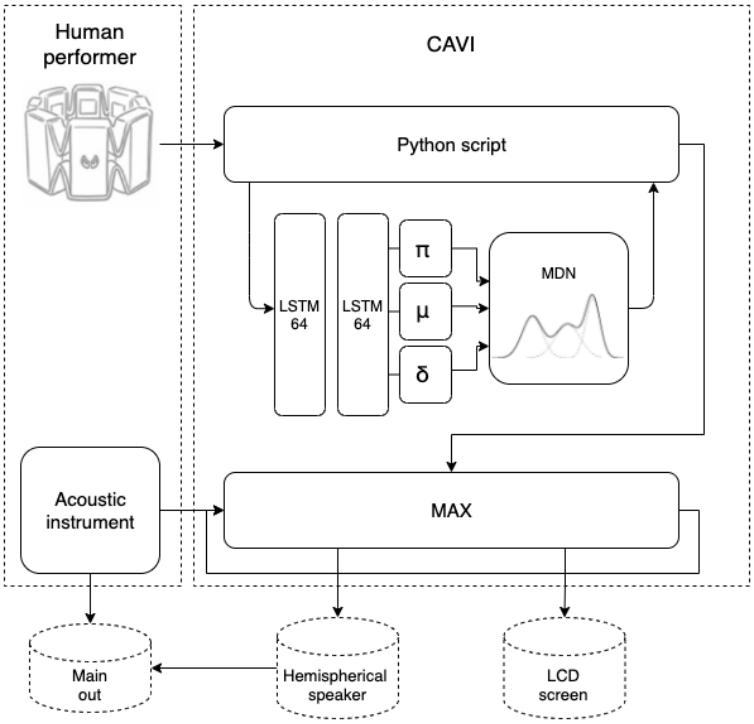


Figure 1. A simplified diagram of CAVI's system architecture: The model receives muscle (EMG) & motion (ACC) data from the Myo armband. The MDRNN outputs the mixture distribution parameters, from which we sample a new window of EMG & ACC data. The generated data is sent to a Max/MSP/Jitter patch that generates visuals and processes the dry acoustic instrument in several effects modules.

conceptually, the model is trained on the user's action repertoire and generates control signals following their actions. This echoes what happens when two people have similar action repertoires. For example, expert basketball players can predict the outcome of a shot more accurately than expert watchers, even before the ball leaves the shooter's hand (Aglioti et al. 2008). According to Knoblich et al. (2011), in such situations, there is a potential for emergent coordination by means of common action representations.

To explore that potential of emergent coordination, we abandoned the call-and-response approach. Instead, CAVI continuously tracks the performer's motion input, consisting of four channels of muscle data and three channels of accelerometer data from the *Myo* armband worn on the right forearm. The generated control signals are then mapped to parameters of the EFX modules. One can imagine that as playing an instrument through some EFX pedals while someone else is tweaking the knobs of the devices.

Sounds

In the current project, CAVI was set up to perform with two acoustic musicians, one guitarist and one percussionist. The idea was that each musician would perform on their acoustic instrument and that CAVI would automate live sound processing of the dry instrument sound. As such, CAVI would act as an advanced effects module. The processing was primarily focused on time-based sound manipulation, such as multiple layers of delay, a spectral time-stretch by Charles (2008), stutter, spectral distortion, and a plate reverb (Dattorro 1997) and space reverb. The audio patch is developed in *Max/MSP*.

The main control interface for the modules is a *live-grid*-based sequencer (see Figure 2). The *jerk* (rate of change of acceleration) of the generated ACC triggers the sequencer to the next step, which functions as a matrix that routes the EFX sends and returns. Every time CAVI executes an action, the bar moves forward or jumps to another

location. The generated EMG, which corresponds to the extension and flexion muscles of the right forearm, controls the EFX parameters. That allows for automating more than a dozen delay lines, their time-based modulations, depths, and send & return values. Considering that modules are interconnected, the fluctuations can grow exponentially, creating extremely dense sound textures. That is where the live-grid becomes handy as the user can draw patterns on a touch screen interface before the performance, edit or call presets during the performance, or entirely randomize these processes.

The central part of the automation is based on a MDRNN running within a custom Python script that fetches data from the Myo armband via Bluetooth, does the preprocessing, windowing, generation, and streaming to Max through Open Sound Control (OSC). The second part in Max relies on real-time analysis modules that track the dry audio input and adjust EFX parameters according to pre-defined thresholds, such as onsets and energy levels. For example, if the performer plays impulsive notes, CAVI increases the reverb time drastically such that it becomes a drone-like continuous sound. Or, if the performer plays loudly, CAVI modifies the dynamic levels based on the performer's quantity of motion (QoM).

Unlike improvisation systems that rely on symbolic music-theoretical data and stylistic constraints, CAVI prioritizes building sound structures in which the performer is expected to navigate spontaneously and even forcefully from time to time. This navigation might be led by a particular sonic event where the performer's and CAVI's actions converge. The performer can focus on a global structure and follow the energy trajectories to influence the textural density. After all, even though CAVI also has "a life of its own" similar to feedback instruments, it is not a fully autonomous agent. Live sound processing is inherently indigent regardless of the controlling agent.

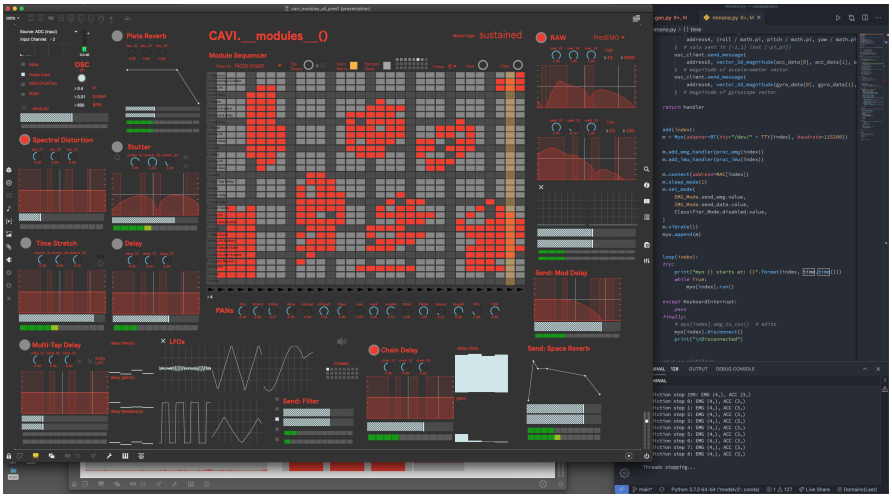


Figure 2. The "cockpit" of CAVI, with the Max patch on the left. The user draws "shapes" on the live grid to determine the overall compositional structure. Each of the EFX modules has individual send/return and is interconnected with other modules. The Python script on the right continuously retrieves data from the Myo armband, pre-processes, feeds into the model, and finally streams the generated data through OSC to Max.

Visuals

CAVI is an audiovisual instrument. We created the "virtual embodiment" of CAVI in *Max/Jitter* using *OpenGL* (see Figure 3). The main motivation behind its visual appearance was to facilitate potential causality ambiguities. The design is based on two layers of *jit.matrix*. The first layer contains digitized pixels of a virtual body shape that is hand-drawn by Katja Henriksen Schia. The second layer encapsulates $350 * 350$ particles on a two-dimensional plane. Initiated with *jit.noise*, these particles are shaped as circles in *jit.gen*. In the same environment, the circled pattern that encapsulates particles is animated as an eye-like shape, attracting the center and the circumference. The generated EMG and ACC data are mapped to the attraction and acceleration parameters of the particles.

We envisioned CAVI as an abstract and cute creature. It is based on a single, large eye, small mouth and tiny legs. As such, it has life-like characteristics, but it is also unnatural.

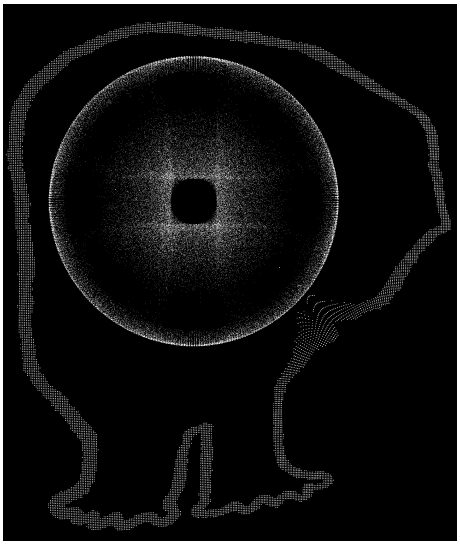


Figure 3. "Say Hi to CAVI. Perhaps you get a blink in return."

It can move but with a limited action repertoire. CAVI can blink when triggering events, open its eye wide when the density of low frequencies increases, or stay calm according to the overall energy levels. We think of it as a toddler that mimicking the parents' gestures, hence the labeling as a *musical AI "toddler"*.

The Performance

We invited two expert musicians from Norway's improvisation scene, Christian Winther (guitar) and Dag Erik Knedal Andersen (drum set). Both excel in free forms of improvised music but are not experienced with interactive music systems. First, they tested CAVI in the *fourMs* Lab at RITMO, University of Oslo (UiO). We provided no training except for a simple introduction to the system. Both test sessions lasted around 30 minutes.

In the ensuing week, they each performed a live duo improvised set with CAVI in a public event titled "Human–Machine Improvisation," which we organized in

Figure 4. Photos from CAVI's first public performance at the Science Library, UiO. CAVI played two sets. The first one was with Christian Winther on guitar. The human musicians were placed to the right on the stage while CAVI's audiovisual output were on the left side of the stage. (Photo: Alena Clim)

collaboration with the Science Library at the University of Oslo. The stage was designed for duo performance (Figures 4 and 5). The right side of the stage was reserved for the human musician while a TV screen and a hemispherical speaker was placed on the left side for CAVI's audiovisual output. We routed audio outputs of the acoustic instrument through a main mixing desk, where dry signals were split between the main out and the computer that runs CAVI's programs. Processed signals were sent back to the main desk to be live-mixed for the sound system and recording. The main output relied on a single floor monitor in front of the human performer to alleviate unwanted acoustic feedback from a PA system.

The Experience

The evaluation of CAVI is based on two datasets. First, we conducted semi-structured one-to-one interviews with each musician. These were recorded, transcribed, and

Figure 5. The second set was with Dag Erik Knedal Andersen on drum set. The human musicians were placed to the right on the stage while CAVI's audiovisual output were on the left side of the stage. (Photo: Alena Clim)

analyzed using a theme-based approach. Second, we asked audience members to fill out a questionnaire immediately following the performance. A total of 20 audience members (7 male, 6 female, 6 other/non-binary, 1 undisclosed, age $M = 34, SD = 7$ years) ranging from music students, professional musicians to avid music listeners took part in the anonymous survey. The audience members' familiarity with electroacoustic improvisation was $M = 7, SD = 3$. We focused on four multiple-choice and linear scale questions in addition to an optional text box for personal comments (see Figures 6 and 7 for the plots).

In the following, we discuss results from both musicians and audience members under two of the pre-defined conceptual dimensions (sense of agency and emergent coordination) in addition to two other concepts (surprise and environment) that emerged through the thematic analysis.

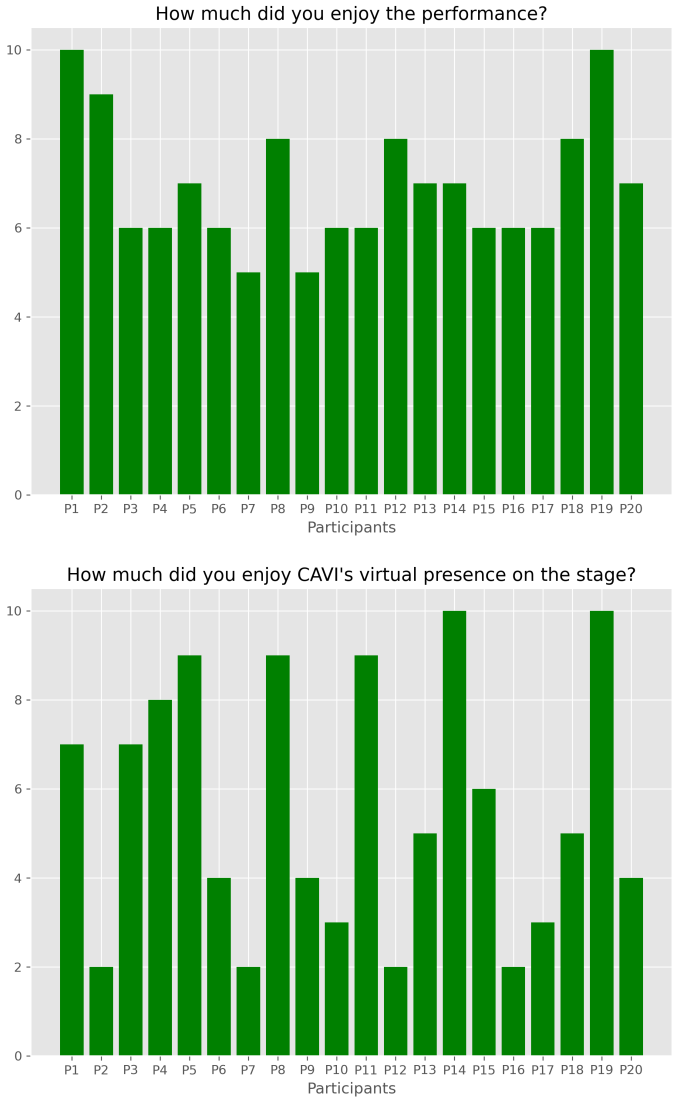


Figure 6. Audience responses regarding their overall enjoyment of the performance (top), and their enjoyment of CAVI's virtual embodiment (bottom).

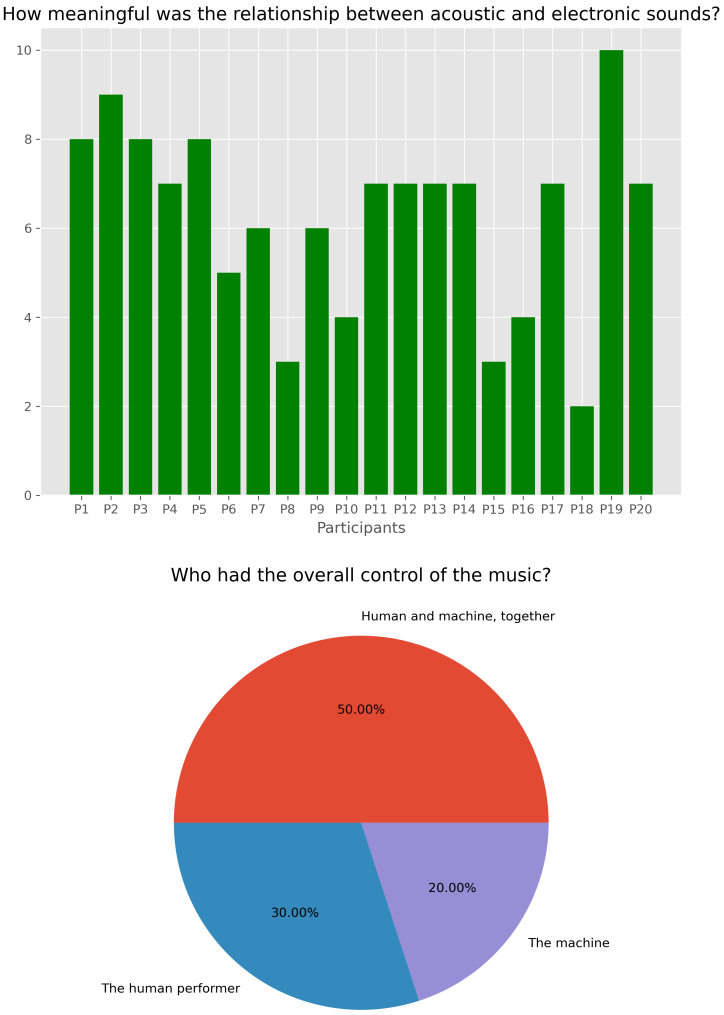


Figure 7. Audience responses regarding how meaningful was the human–CAVI relationship (top), and their thoughts about the overall control of the music (bottom).

Sense of Agency

In the interview with the drummer, he quickly pointed to the sense of agency: "AI pulls you back to the point in time you were previously while now you are ahead of that point," he remarked, "somehow it didn't feel natural to me." Traditionally, most musical tools, such as effects processing, have been designed so that the musician has complete control. Thus the predictability of the system is high. When describing his experience of performing with CAVI he said:

> I would put it somewhere between an effects box and a performer. It's not an effects box but it's also not a performer. It's a gradual transition between those two, in this case, I think.

The guitarist followed up in his interview by referring to the automated effects processing as "a non-pleasant in-betweenness." First and foremost, he was confused by someone else being in control of "his" sound. Not being experienced with such interactive systems, he indicated that he prefers the pure sound of his instrument. "I'm not into layered effects that much," he remarked, and the ambiguity that effects processing creates in general, such as camouflaging small details of the acoustic instrument, is an apt reason for that. He prefers that an improvisation partner brings musical originality so that he can create separate layers with his partner. With CAVI, he felt that they were not really working together. This was also a comment by one of the audience members, who wrote: "they're both just super loud and fighting, not sharing."

The interviews revealed that the initial description of the instrument is essential for how the musicians would approach the performance. "If it were a piece, I would listen to it more," the guitarist stressed, "I would not bring myself into it as much as I did; I would restrain myself more of my output and would kinda surf along [its] output." For later experiments, it may be relevant to think about introducing CAVI as a what Schnell and

Battier (2002) calls a "composed instrument." As also Dahlstedt (2021) reflects on, then there would be the composer's agency to consider.

Emergent Coordination

According to the guitarist, CAVI was like a beginner improviser who used all the chops at once, sounding exhaustingly busy and becoming less surprising over time. He stressed, "if a musical partner only manipulates, you can never have a dialogue." Similarly, one of the audience members also commented:

> I sometimes felt the relationship between musicians and CAVI was combative more than symbiotic, almost like improvisation partners who haven't practiced together much yet.

The drummer commented: Regarding what musicians can expect from such systems, "you think about Arnold Schwarzenegger's Terminator when you think about AI, then you get to a situation, which is very different than you thought it would be." The drummer compared the experience to playing chess with an AI. However, "as opposed to chess, improvisation is very complex and difficult to say 'wow that was a good move' or 'that was a horrible thing to do'."

Normally, joint actions emerge through a complex multimodal perception–action mechanism. In a duo setting where one of the agents is not competent enough, the other agent needs to compromise to achieve musical coordination. According to the guitarist, balanced skills and autonomy is necessary for emergent coordination between partners:

> As an improviser, it is plating with too much information. It is a too big package in a sort of predictable way, which I did not like while playing with it. Because, you bring so many sounds to the table but if I stop, it stops.

The guitarist found that it was necessary to use tacet passages to create "unbalanced" situations. This allowed for more interesting things to emerge.

The drummer mentioned how he dealt with temporal ambiguity by "stretching things" that he played so that it yielded "a musical sense." He explained his mindset by referring to how he had seen the guitarist perform with CAVI:

> I tried to respond to things that happen through the AI, which, I thought, was a better decision than [the guitarist] who was doing his thing, playing the guitar, making mellow sounds, not very interactive. He did not try to communicate with the whole thing. I think you have to let go of the control of it when playing with something like that. It takes you into some uncharted territories, and you have to just go along with it. Because if you try to get it back in the direction you want it to go, it doesn't work. [...] Considering that I and [the guitarist] did two very different things, just listening back to it, [CAVI] adapted our playing.

Uncertainty and Surprise

Although we did not ask about it specifically, three audience members' had similar comments: "the AI lacked the unpredictability of a human performer." Both musicians also commented on the restrictions imposed by a too predictable improvising partner. At one point during the performance, the guitarist expressed his frustration with the predictable nature of CAVI by bending onto the guitar, almost like hugging it. The aim was to mute the guitar and provoke another reaction.

Both the sense of freedom and the level of engagement tend to emerge from appropriate and moderate ways of introducing surprise and complexity (Borgo 2005). The drummer explained how he thinks about developing ideas in improvised performances:

(1) Continue with an existing idea; (2) surprise; (3) stop. This is an efficient way of quickly developing more complex musical content: "every choice you make will affect the music and the performance and make other players make other choices."

"I want it to be more machine, in a way almost like dead," the guitarist remarked by referring to John Cage's *chance operations*. At the same time, he commented how expert improvisers actively learn to perform with others. What you know from playing with a musician can influence how you play with someone else: "it would be great if [CAVI] could learn something from the drummer's session." That is an interesting point as the concept of memory is closely linked with the idea of familiarity. Hence, in this context, memory might imply an attraction to random or surprising elements. However, CAVI's model was trained on a fixed dataset and did not feature online.

Environment

In the current performances with CAVI, we optimized the system to capture the electroacoustic guitar sound using a close-proximity microphone. We also wanted to present CAVI through a single-point hemispherical speaker and TV on stage. That created the sense of one human and one machine performer on stage. Unfortunately, the signal levels were adjusted to what turned out to be an unbalanced on-stage listening condition for the performers. Listening back at the recording, they experienced a completely different performance than on stage. "Because of the monitoring/listening situation, I could not catch up with its initiatives. But when I listened to it, I could catch much more. So I was happy that it brought to the table more than I experienced playing with it," the guitarist remarked. The drummer also commented on the difference between performer and listener experiences:

> I think, on my part, it was more about the sound difficulties than playing with [CAVI]. When I listened back to the performance, I've changed my view on the

whole thing quite radically.

Even though the stage setup had been challenging, the guitarist liked the idea of using a hemispherical speaker over a PA. Hemispherical speakers provide a more individual sound source. He remarked he likes when CAVI is "more pointed in its output so that I can better grasp what it is doing."

The drummer used a relatively simple instrument (a snare drum, hi-hat, and a crash) with paired, closely-placed small-diaphragm microphones in the lab rehearsals with CAVI. However, during the performance he used a complete drum set of mediocre quality in concert. "It was a 'dead' sounding drum set," he described. The main disadvantage of such a live setup was the inefficient capture of small details, providing the drummer with a cognitive load due to an unknown instrument with more parts than the smaller set he practiced during the rehearsal.

Both musicians commented that they enjoyed the rehearsals more than the performance. "For example, in the room where we had the practice session, it felt the whole thing was close to me, whereas when playing the concert it felt like it was very distant," the drummer said. He would have preferred a small, dry jazz club over the fairly large, lively concert location of the current performance. In such a setting, he would only use proximity and contact microphones on the drum set. "A small space would certainly help my thing, like getting more in touch feeling a bit more of the vibe and the whole thing." Despite the sound problems, he too agreed that sitting close to a TV and hemispherical speaker was a nice setup for such a performance. Just like the guitarist, he was positively surprised when listening to the recording: "But then the recording was full-on, the whole thing, the whole time! And that was very nice," he concluded.

Discussion

Most traditional acoustic instruments can be explored within different styles and genres and be used to develop specialized performance techniques. On the other hand, as Magnusson (2009) suggests, computational systems require concepts and tailored programming languages for a more or less pre-composed interactive scenario and sonic output. CAVI is something in between. It has instrumental qualities, but it more resembles an advanced effects module. However, the development process also included compositional thinking. In many ways, the drummer summed it up nicely:

> Thinking about the whole thing, it's kind of conceptual composition or a conceptual work of art.

Both musicians' and some of the audience members' feedback stressed the lack of creative unpredictabilities. Surprising elements are essential to the positive aesthetic experience of improvised music, and they should be a critical consideration when designing such systems. However, unlike the Cagean approach, the notion of surprise in improvised music is different from the noise of a random number generator. The lack of familiarity can be perceived as impotence and discrepancy in collaborative decision-making.

Despite several shortcomings, the listening experience was appreciated by both musicians and the majority of the audience members. However, this is not necessarily synonymous with whether CAVI was successful or not as an improvisation partner. One limitation of CAVI is the lack of feelings and a sense of aesthetics. As opposed to a human that listens and adjusts, CAVI perseveres with its agenda. Interestingly, the guitarist and drummer chose two different strategies in their interaction. The guitarist insisted on his intents while the drummer preferred making compromises and allowed coordination to emerge with the "toddler."

Finally, both musicians strongly stressed the difference between the experience of playing and listening back to it. As Schiavio (2015) argues, the environment actively co-constitutes music together with the living bodies and their activities. The performance space, microphone setup and monitoring system are all parts of the dynamicity of emergence, control, and agency. Then, the room and technical rig become critical tools and decisions concerning the musical composition and contribute to the musical agency assigned to or shared with the artificial musical agent.

In sum, our experiences with CAVI thus far can be summarized as follows:

1. The sense of agency, hence whether the system is experienced as a tool or an actor, varies throughout a performance and strongly depends on the perceived qualities of the collaborative performance. These qualities, however, vary from one musician to another, or, from whether the it is a composed piece or free improvisation.
2. Surprise is not only an important aesthetic component in improvised music but can also compensate the ambiguity in the sense of agency stemming from joint actions.
3. Collaborative improvisation is a highly multimodal practice. Therefore, environmental factors, such as the physical space and acoustics, stage design, and technical rigging, are crucial for the performance quality, which is closely linked with the agency experience.

CAVI started life as a baby and is currently at the level of a toddler. The aim is to continue to build on its multimodal factors and emerging sense of agency. The aim is to move towards a more intuitive collaborative human–machine system for music performance.

References

Aglioti, S. M., P. Cesari, M. Romani, and C. Urgesi. 2008. "Action anticipation and motor resonance in elite basketball players." *Nature Neuroscience* 11(9):1109–1116. URL https://www.nature.com/articles/nn.2182. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 9 Primary_atype: Research Publisher: Nature Publishing Group.

Bailey, D. 1993. *Improvisation: its nature and practice in music*. New York: Da Capo Press.

Borgo, D. 2005. "Rivers of Consciousness: The Nonlinear Dynamics of Free Jazz." In *Jazz Research Proceedings Yearbook*, pp. 46–58. URL https://www.academia.edu/1337717/Rivers_of_Consciousness_The_Nonlinear_Dynamics_of_Free_Jazz.

Charles, J.-F. 2008. "A Tutorial on Spectral Sound Processing Using Max/MSP and Jitter." *Computer Music Journal* 32(3):87–102. URL https://www.mitpressjournals.org/doi/abs/10.1162/comj.2008.32.3.87.

Collins, N. M. 2006. "Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems." PhD dissertation, University of Cambridge. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.2661&rep=rep1&type=pdf.

Dahlstedt, P. 2021. "Musicking with Algorithms: Thoughts on Artificial Intelligence, Creativity, and Agency." In E. R. Miranda, ed. *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer International Publishing, pp. 873–914. URL https://doi.org/10.1007/978-3-030-72116-9_31.

Dattorro, J. C. 1997. "Effect Design: Part 1: Reverberator and Other Filters." *Journal of Audio Engineering Society* 45(9):660–684. URL https://ccrma.stanford.edu/~dattorro/EffectDesignPart1.pdf.

Ebert, J. P., and D. M. Wegner. 2010. "Time warp: Authorship shapes the perceived timing of actions and events." *Consciousness and Cognition* 19(1):481–489. URL https://www.sciencedirect.com/science/article/pii/S1053810009001548.

Ellefsen, K. O., C. P. Martin, and J. Torresen. 2019. "How do Mixture Density RNNs Predict the Future?" *arXiv:1901.07859 [cs, stat]* URL http://arxiv.org/abs/1901.07859. ArXiv: 1901.07859.

Erdem, C., Q. Lan, and A. R. Jensenius. 2020. "Exploring relationships between effort, motion, and sound in new musical instruments." *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* 16(3):310–347. URL https://humantechnology.jyu.fi/archive/vol-16/issue-3/erdem_lan_jensenius.

Fels, S. S., L. Kaastra, S. Takahashi, and G. Mccaig. 2004. "Evolving Tooka: from Experiment to Instrument." pp. 1–6. URL https://zenodo.org/record/1176595.

Gentsch, A., and S. Schutz-Bosbach. 2015. "Agency and Outcome Prediction." In P. Haggard and B. Eitam, eds. *The Sense of Agency*. Oxford University Press, pp. 217–234. URL https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190267278.001.0001/acprof-9780190267278-chapter-9.

Graves, A. 2013. "Generating Sequences With Recurrent Neural Networks." *arXiv:1308.0850 [cs]* URL http://arxiv.org/abs/1308.0850. ArXiv: 1308.0850.

Ha, D., and D. Eck. 2017. "A Neural Representation of Sketch Drawings." *arXiv:1704.03477 [cs, stat]* URL http://arxiv.org/abs/1704.03477. ArXiv: 1704.03477.

Haggard, P. 2005. "Conscious intention and motor cognition." *Trends in Cognitive Sciences* 9(6):290–295. URL https://linkinghub.elsevier.com/retrieve/pii/S1364661305001191.

Haggard, P., S. Clark, and J. Kalogeras. 2002. "Voluntary action and conscious awareness." *Nature Neuroscience* 5(4):382–385. URL http://www.nature.com/articles/nn827.

Bandiera_abtest: a Cg_type: Nature Research Journals Number: 4 Primary_atype: Research Publisher: Nature Publishing Group.

Jeannerod, M. 2008. "The sense of agency and its disturbances in schizophrenia: a reappraisal." *Experimental Brain Research* 192(3):527. URL https://doi.org/10.1007/s00221-008-1533-3.

Kaltenbrunner, M., S. Jorda, G. Geiger, and M. Alonso. 2006. "The reacTable*: A Collaborative Musical Instrument." In *15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06)*, pp. 406–411. URL http://ieeexplore.ieee.org/document/4092244/.

Kawabe, T. 2013. "Inferring sense of agency from the quantitative aspect of action outcome." *Consciousness and Cognition* 22(2):407–412. URL https://www.sciencedirect.com/science/article/pii/S1053810013000093.

Kiefer, C., D. Overholt, and A. Eldridge. 2020. "Shaping the behaviour of feedback instruments with complexity-controlled gain dynamics." URL https://zenodo.org/record/4813406. Conference Name: International Conference on New Interfaces for Musical Expression Pages: 343-348 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv:1412.6980 [cs]* URL http://arxiv.org/abs/1412.6980. ArXiv: 1412.6980.

Knoblich, G., S. Butterfill, and N. Sebanz. 2011. "Psychological Research on Joint Action." In *Psychology of Learning and Motivation*, vol. 54. Elsevier, pp. 59–101. URL https://linkinghub.elsevier.com/retrieve/pii/B9780123855275000036.

Knoblich, G., and B. H. Repp. 2009. "Inferring agency from sound." *Cognition* 111(2):248–262.

Kosowitz, S., and L. Vickery. 2013. "Retaining a sense of spontaneity in Free Jazz improvisation through music technology." *Research outputs 2013* URL https://ro.ecu.edu.au/ecuworks2013/273.

Launay, J. 2015. "Musical Sounds, Motor Resonance, and Detectable Agency." *Empirical Musicology Review* 10(1-2):30–40. URL https://emusicology.org/article/view/4579. Number: 1-2.

Lee, M., A. Freed, and D. Wessel. 1991. "Real-Time Neural Network Processing of Gestural and Acoustic Signals." pp. 277–280.

Lewis, G. E. 2000. "Too Many Notes: Computers, Complexity and Culture in Voyager." *Leonardo Music Journal* 10(1):33–39. URL https://muse.jhu.edu/article/20320.

Liontiris, T. P. 2018. "Low Frequency Feedback Drones: A non-invasive augmentation of the double bass." URL https://zenodo.org/record/1302605. Conference Name: International Conference on New Interfaces for Musical Expression Pages: 340-341 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

Magnusson, T. 2009. "Epistemic tools: the phenomenology of digital musical instruments." doctoral, University of Sussex. URL http://sro.sussex.ac.uk/id/eprint/83540/.

Martin, C. P. 2019. "IMPS: Interactive Musical Prediction System: Demo Video." URL https://zenodo.org/record/2597494.

Martin, C. P., and J. Torresen. 2019. "An Interactive Musical Prediction System with Mixture Density Recurrent Neural Networks." URL https://zenodo.org/record/3672952. Conference Name: International Conference on New Interfaces for Musical Expression Pages: 260-265 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

Melbye, A. P., and H. A. Ulfarsson. 2020. "Sculpting the behaviour of the Feedback-Actuated Augmented Bass: Design strategies for subtle manipulations of string feedback using simple adaptive algorithms." URL https://zenodo.org/record/4813328. Conference Name: International Conference on New Interfaces for Musical Expression Pages: 221-226 Publication Title: Proceedings of the International Conference on New Interfaces for Musical Expression Publisher: Zenodo.

Mendoza, J. I., and M. R. Thompson. 2017. "Gestural Agency in HumanâMachine Musical Interaction." In M. Lesaffre, P.-J. Maes, and M. Leman, eds. *The Routledge Companion to Embodied Music Interaction*. New York ; London : Routledge, 2017.: Routledge, 1st edition, pp. 412–419. URL https://www.taylorfrancis.com/books/9781317219736/chapters/10.4324/9781315621364-45.

Minsky, M. 1981. "Music, Mind, and Meaning." *Computer Music Journal* 5(3):28–44. URL http://www.jstor.org/stable/3679983. Publisher: The MIT Press.

Moore, J. W. 2016. "What Is the Sense of Agency and Why Does it Matter?" *Frontiers in Psychology* 7:1272. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5002400/.

Nort, D. V., P. Oliveros, and J. Braasch. 2013. "Electro/Acoustic Improvisation and Deeply Listening Machines." *Journal of New Music Research* 42(4):303–324. URL https://doi.org/10.1080/09298215.2013.860465.

Pachet, F. 2003. "The Continuator: Musical Interaction With Style." *Journal of New Music Research* 32(3):333–341. URL https://www.tandfonline.com/doi/abs/10.1076/jnmr.32.3.333.16861. Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1076/jnmr.32.3.333.16861.

Russell, J. A. 1980. "A circumplex model of affect." *Journal of Personality and Social Psychology* 39(6):1161–1178. Place: US Publisher: American Psychological Association.

Russell, S. J. S. J. 2010. *Artificial intelligence : a modern approach*. Third edition. Upper Saddle River, N.J. : Prentice Hall. URL https://search.library.wisc.edu/catalog/9910082172502121.

Sawyer, R. K., and S. DeZutter. 2009. "Distributed creativity: How collective creations emerge from collaboration." *Psychology of Aesthetics, Creativity, and the Arts* 3(2):81–92. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/a0013282.

Schiavio, A. 2015. "Action, Enaction, Inter(en)action." *Empirical Musicology Review* 9(3-4):254–262. URL https://emusicology.org/article/view/4440. Number: 3-4.

Schmidhuber, J. 2009. "Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes." *arXiv:0812.4360 [cs]* URL http://arxiv.org/abs/0812.4360. ArXiv: 0812.4360.

Schnell, N., and M. Battier. 2002. "Introducing Composed Instruments, Technical and Musicological Implications." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 156–160. URL https://zenodo.org/record/1176460.

Schuster, M. 1999. "On Supervised Learning From Sequential Data With Applications For Speech Recognition." Ph.D. Thesis, Nara Institute of Science and Technology, Nara, Ikoma, Japan. URL https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.17.1460&rep=rep1&type=pdf.

Tahiroglu, K., M. Kastemaa, and O. Koli. 2021. "AI-terity 2.0: An Autonomous NIME Featuring GANSpaceSynth Deep Learning Model." In *International Conference on New Interfaces for Musical Expression*. URL https://nime.pubpub.org/pub/9zu49nu5/release/1.

Tanaka, A., and M. Donnarumma. 2018. "The Body as Musical Instrument." *The Oxford Handbook of Music and the Body* URL http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780190636234.001.0001/oxfordhb-9780190636234-e-2.

Tatar, K., and P. Pasquier. 2017. "MASOM: A Musical Agent Architecture based on Self-Organizing Maps, Affective Computing, and Variable Markov Models." p. 8. URL https://zenodo.org/record/4285244.

Weinberg, G., and S.-L. Gan. 2001. "The Squeezables: Toward an Expressive and Interdependent Multi-Player Musical Instrument." *Computer Music Journal* 25(2):37–45. URL https://www.jstor.org/stable/3681530. Publisher: The MIT Press.

van der Wel, R. P., N. Sebanz, and G. Knoblich. 2012. "The sense of agency during skill learning in individuals and dyads." *Consciousness and Cognition* 21(3):1267–1279. URL https://linkinghub.elsevier.com/retrieve/pii/S1053810012000736.

Wolpert, D. M., Z. Ghahramani, and M. I. Jordan. 1995. "An Internal Model for Sensorimotor Integration." *Science* 269(5232):1880–1882. URL http://www.jstor.org/stable/2889276. Publisher: American Association for the Advancement of Science.

Paper VI

What Makes Interactive Art Engaging?

Michael Krzyzaniak, Çağrı Erdem, Kyrre Glette

Published in *Frontiers in Computer Science*, 4.
April 2022

frontiers | Frontiers in Computer Science

ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fcomp.2022.859496

What Makes Interactive Art Engaging?

Michael Krzyzaniak, Çağri Erdem and Kyrre Glette*

RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway

Interactive art requires people to engage with it, and some works of interactive art are more intrinsically engaging than others. This article asks what properties of a work of interactive art promote engagement. More specifically, it examines four properties: (1) the number of controllable parameters in the interaction, (2) the use of fantasy in the work, (3) the timescale on which the work responds, and (4) the amount agency ascribed to the work. Each of these is hypothesized to promote engagement, and each hypothesis is tested with a controlled user study in an ecologically valid setting on the Internet. In these studies, we found that more controllable parameters increases engagement; the use of fantasy increases engagement for some users and not others; the timescale surprisingly has no significant on engagement but may relate to the style of interaction; and more ascribed agency is correlated with greater engagement although the direction of causation is not known. This is not intended to be an exhaustive list of all properties that may promote engagement, but rather a starting point for more studies of this kind.

Keywords: interactive art, fun, engagement, web-based interaction, user studies

OPEN ACCESS

Edited by:
Gerrit C. Van Der Veer,
University of Twente, Netherlands

Reviewed by:
Danzhu Li,
University of Twente, Netherlands
Bert Bongers,
University of Technology Sydney, Australia

****Correspondence:***
Michael Krzyzaniak
mkrzyzaniak@protonmail.com

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *21 January 2022*
Accepted: *22 March 2022*
Published: *26 April 2022*

Citation:
Krzyzaniak M, Erdem Ç and Glette K (2022) What Makes Interactive Art Engaging? Front. Comput. Sci. 4:859496. doi: 10.3389/fcomp.2022.859496

1. INTRODUCTION

Interactive art is art that you can play with. It responds to the actions of its interactants.[1] Such works are typically either visual or sonic in nature, involve digital technology, and respond to the movements, sounds, or input (*via* a computer interface) of the interactant. This creates a bidirectional flow of information between the interactant and the work. The interactant's actions are, therefore, an integral part of interactive art; the proverbial tree falling in the forest definitely does not make any sound in the absence of observers, if it depends on someone being there to fell it in the first place.

Consequently, the idea of *engagement* underlies all interactive art. In order for a work to be complete, an observer has to be sufficiently engaged so as to voluntarily perform the actions to which the work responds. This gives rise to the overall question of this article:

What properties should a work of interactive art have in order to promote engagement?

Stated another way, how can these works be designed to be fun, so that people want to interact with them? Engagement may be operationalized as the amount of time that people spend voluntarily interacting with such works. So how can a work be designed to maximize the amount of time people spend interacting with it?

The amount of time people spend looking at art in general has been studied. A seminal study in Smith and Smith (2001) found that museum visitors spent 27.2 s on average (with a median

[1] I will use the term "interactant" throughout this article to refer to a human who engages with a work of interactive art.

of 17.0 s) looking at individual paintings, including the time spent reading the accompanying label. A larger 2017 followup study replicated these findings (Smith et al., 2017), with no significant differences as compared to the first study. The followup, which was conducted after the invention of smartphones, additionally found that some visitors took selfies with paintings without actually viewing the paintings, which at least suggests that the presence of digital technology can change how people engage with art. In both studies, the authors observe that some paintings were viewed for significantly longer than others. However, they do not examine whether there are intrinsic properties of the paintings that may account for this, although they do note that it may relate to the presence or absence of seating near the painting.

Engaging properties of other types of systems have been studied. Seminal studies by Malone (1981) investigated this question with regard to educational computer games for children. The studies found that to promote engagement, games should have a goal with uncertain outcomes, should make use of fantasy, and should promote curiosity *via* an optimal level of information complexity. Games, however, differ from interactive art in the following way: Games by definition have fixed goals where players try to achieve something specific that is known beforehand. Interactive art, by contrast, either has no goals, or emerging ones, and interactants are supposed to interact for the sheer moment-to-moment pleasure of doing so. Consequently, it is not clear how well these principles translate to interactive art, although further analysis is presented in section 3 below.

Since then, a healthy literature has emerged on fun and enjoyment in computer systems. A considerable amount of this work is compiled in the 2002 book Funology (Monk et al., 2002), and its 2018 followup Funology 2 (Blythe and Monk, 2018). These contain studies on computer games (Pagulayan et al., 2003), dating apps (Zytko et al., 2018), information displays (Ljungblad et al., 2003), and other types of computer systems. Dating apps and information displays are *tools* in the sense that people use them in order to accomplish something, whereas interactive artworks are *toys* in the sense that there is no external reason to use them. Tools undoubtedly promote engagement differently than toys. Regarding toys, in Sykes and Wiseman (2003), the authors argue that fear is fun, and they demonstrate this by presenting a "haunted" VR experience at a science festival. Similarly, in Fernaeus et al. (2018), the authors posit that bodily movement promotes enjoyment, and they support this by presenting several systems that they designed to illustrate the point. These include interactive artworks, for example a lamp that follows your breathing. However, neither paper presents a controlled experiment that shows that people actually enjoy fear or movement more than some baseline systems. In fact, out of the 38 articles on how to design fun and engaging computer systems in Blythe and Monk (2018), many of them, for example (Overbeeke et al., 2003), contain very specific opinions about what properties of a system promote engagement; yet only three or four of them (Karat et al., 2002; Desmet, 2003; Pagulayan et al., 2003; Rosson and Carroll, 2018) substantiate those opinions with a controlled quantitative experiment similar to the Malone studies, and those are not about interactive art.

The artist Brigid Costello compiled a comprehensive theoretically grounded list of properties that make interactive art pleasurable (Costello and Edmonds, 2007). The list contains, e.g., creation, exploration, discovery, difficulty, et cetera. She designed a new work called *Just a bit of Spin* to make use of these properties, and showed it in a museum. However, she noted that although visitors explored the work, they did not *play* with it. In a followup study (Costello and Edmonds, 2009), she hypothesized that this was due to the work's low complexity, although complexity was not on the original list of properties. After redesigning the work to be more complex, she found that museum visitors did spend more time interacting with it as compared to the original version. In Bongers and Mery (2011), displayed an interactive artwork in a museum and collected participant data. They found that visitors spent about a minute on average interacting with the artwork. The visitors spent a portion of this time engaging in behaviors that were not designed parts of the interaction. The authors in particular note social behaviors, like the visitors explaining the work to one another, and arguing with one another over the use of the interfaces that control the work.

For the sake of completeness, it is worth pointing out that the perverse way to maximize the amount of time people spend interacting with digital systems is to get them addicted by exploiting human psychology. This technique has been highly optimized by both the video game and social media industries, which have an incentive of hundreds of billions of dollars annually[2,3] to encourage addiction. For example, the use of rewards to maximize dopamine production is a well researched topic (Sapolsky, 2017) that is often exploited in games, e.g., through the use of gradually diminishing rewards.[4] Likewise, social media sites actively remove cues that users would use to monitor their own usage, for example through the use of infinite scroll (Chou et al., 2005). Although similar techniques could undoubtedly be applied to interactive art, seeking to addict a user is different than seeking to engage them, even if these are both operationalized by duration of interaction. The difference is that in an engaging system, the user spends time for their own benefit, for their own leisure or edification, while in an addicting system, they spend their time for someone else's benefit and even to their own detriment, e.g. because their time is being monetized by a corporation. So while it is well studied how to addict people, it is less well known how to engage them in a healthy and edifying context such as is provided by art.

In light of the foregoing observations, the present paper provides a starting point for understanding how certain properties of an interactive artwork relate to the way an interactant voluntarily engages with it. Four separate studies are presented herein, each examining a different property. The first study pertains to the number of controllable parameters of a work of interactive art; the second to the use of fantasy in the work; the third to the timescales on which the work responds

[2]https://www.grandviewresearch.com/industry-analysis/video-game-market
[3]https://www.ibisworld.com/industry-statistics/market-size/social-networking-sites-united-states/
[4]https://levelskip.com/how-to/Skinners-Box-and-Video-Games

to input; and the fourth to the amount of agency an interactant ascribes to the work. This is not intended to be an exhaustive list of properties that might promote engagement, and are just a few of the properties that the authors have observed to be present in varying degrees in real work of the genre. The studies were conducted by posting bespoke interactive artworks on the internet where visitors were able to interact with them in an ecologically valid setting. This technique, which will be described in greater detail anon, has been fruitful and could be used to explore other properties in the future.

2. STUDY 1—NUMBER OF CONTROLLABLE PARAMETERS

Different works of interactive art have different numbers of controllable parameters, where a degree of freedom is a parameter that the visitor can adjust. The work of Brigid Costello discussed in the introduction illustrates this clearly. The piece consists of a disk that interactants can spin to play recorded sounds. The original version has two controllable parameters; the direction of spin selects which recordings will be played back, and the speed of spin controls the speed of audio playback. The second version of the work introduced a "scratching" gesture that allowed interactants to cycle through different sets of recordings, providing an additional degree of freedom. As another example, consider tabletop user interfaces. *Sandscape* by the Tangible Media group at MIT (Ishii et al., 2004), in its most well-known form, is a sandbox with a heightmap of the sand projected onto it from above.This effectively has one macroscopic degree of freedom; the height of the sand controls the color of the projection. By contrast, *Reactable* by the Music Technology Group at UPF (Jordà et al., 2005) has many controllable parameters. Users create sound by placing fiducial markers on a table. A marker's type, location, orientation, and distance to other markers can control the waveform, frequency, amplitude, and other properties of the sound. Some markers can modify the sounds of other markers, e.g., *via* frequency modulation or filtering, with the relevant parameters also controllable. This results in a large number of controllable parameters. This raises the research question for Study 1:

Do users engage longer with interactive artworks that have more controllable parameters?

2.1. Design

To test this question, I designed the widget shown in **Figure 1**. The widget was created using common web technologies and runs in any modern web browser at the time of writing. It consists of a canvas that displays a procedurally-drawn animation, two buttons, and a bank of sliders. At each frame of animation, a new ellipse is drawn on the canvas. The hue, rotation angle, and location of the ellipses vary over time, with the ellipse locations broadly wandering around the canvas following a Lissajous curve. The sliders allow visitors to adjust the animation parameters, the size of the ellipse, the speed at which it progresses around the canvas, and so forth. Additionally, if a visitor clicks (or touches) the canvas, the ellipse locations will orbit the cursor (or finger) instead of following the Lissajous curve, and will be drawn in grayscale instead of color. Of the two buttons, one allows visitors to clear the canvas, making it entirely white, and another that allows visitors to save the canvas as it currently appears to their computers as a regular image file.

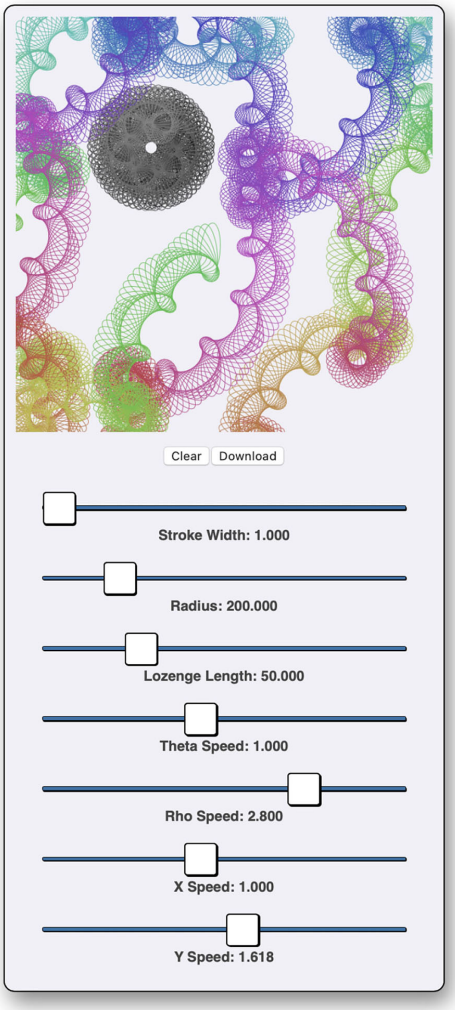


FIGURE 1 | The widget used in Study 1.

Participants in the study were assigned randomly to one of two conditions, called *sliders* and *no-sliders*. Visitors in the *sliders* condition were presented the interface exactly as it is shown in **Figure 1**. Visitors in the *no-sliders* condition were presented

an identical interface, except the sliders were hidden and the associated parameters could not be adjusted, representing a reduced number of controllable parameters. I kept track of each web browser that visited the page, so if the same browser visited more than once, it would be presented the same condition each time. The widget is available for reference on the internet, and the individual conditions can be accessed *via* the following URLs.

1. redacted_for_anon_review
2. redacted_for_anon_review

2.2. Data Collection

I posted this widget to my biography page on the University of Oslo website. I removed all other content from the page, except for the standard navigational elements belonging to the enclosing page template. I recorded the amount of time each visitor spent on the page, along with other standard analytics data, which I describe in more detail in section 2.3, below. I did not collect any personally identifying information nor IP addresses. All visitors to this page had already consented to the university's cookie policy, which covers the collection of non-identifying analytics and usage-pattern data. This provided the most natural and ecologically valid setting for the study. I recruited participants first by sending a hyperlink to a small mailing list of a limited number of my colleagues, alerting them that I had made a fun diversion for them to play with during the 2020 university closure, which was in effect at the time the study was conducted. Subsequently I included a prominent hyperlink to my bio page at the bottom of all emails that I sent to anyone. Over time this was a reliable way of recruiting participants.

2.3. Data Preprocessing

Because the study was conducted "in the wild," the data are somewhat messier than they would be in a laboratory study, and consequently I was obligated to make decisions about how they should be filtered. In this study, I applied the following preprocessing steps to the data in exactly this order:

1. I monitored the user-agent string for search indexing bots. No data was collected from a bot that declared itself as such, although it is likely that some bots can and do execute javascript and simulate input events. I was not able to collect IP addresses because they are personally-identifying, and consequently I was not able to check against lists of known bots. Nonetheless, I do not believe that any data was collected from bots.
2. Some of the researchers associated with the study may have had unrelated reasons to visit my biography page during data collection. In order to exclude their data from the study while maintaining anonymity for all visitors, these researchers were given a special URL. When they visited the URL, the server created a record in the database that marked their browser as belonging to a "developer." This record allowed all previous and future visits from that browser to be excluded from all studies in this article.
3. I measured the number of seconds each visitor spent on the page, from the time it loaded until they navigated away. From that I subtracted out any period of time when the window was not in focus, e.g., because the visitor had another tab or a different application in the foreground.
4. Because some visitors might have opened the page and left it in focus while wandering off to prepare a sandwich, I also monitored input events on the page, such as moving the mouse over the page, clicking, scrolling, and touching the page. I subtracted out any period of inactivity greater than 10 s in which no input events occurred. I will refer to the amount of time left after making these subtractions as the "active" time the visitor spent on the page.
5. If a browser visited the page within 10 s of having navigated away from it, e.g., because the visitor refreshed the page, I appended the new visit to the previous visit, treating both as a single visit, with the period between visits treated as though the page were not in focus.
6. Some visitors spent an implausibly short period of time on the page, with two visitors spending only 2 s each. These visits were consistent with browsers pre-loading the page in the background without the visitor ever actually navigating to the page. Moreover, because the animation started automatically on page load, real visitors could enjoy it without clicking on anything or performing other trackable activities. This was a flaw in the study design that meant that for very short visits in particular, it was in some cases impossible to determine whether the page was actually displayed to the visitor. Consequently, I removed all visits that were less than 20 active seconds in duration, which removed the ambiguous cases. The remaining studies in this paper corrected this design flaw, by making visitors perform some action that proves that they interacted with the widget.
7. Some browsers visited the page more than once, e.g., on different days. In the canonical version of this study I only included the first visit from each visitor, so that individual visitors would not have disproportionate influence on the results, and because experienced visitors might interact differently than first-time visitors. As a special case I will also present some analysis on the number of visits per browser, but unless explicitly stated, I only include the first visit per browser.

In total, 28 browsers not belonging to known bots or developers visited the page a total of 44 times during the data collection period, resulting in 31 min and 13 s of active page time. After preprocessing, there were 22 remaining participants, with one visit by each included, totaling 21 min and 42 s of active page time. Ten of these were randomly assigned to *no sliders*, and 12 to *sliders*. Only two of these were on touch input devices, one tablet and one mobile phone, both assigned to the *sliders* condition, while the remainder were all traditional cursor input devices.

2.4. Results

2.4.1. Did the Participants That Were Presented Extra Controllable Parameters Explore Them?

Two out of 12 visitors in the *sliders* group did not move any of the sliders, although both of them did click the canvas. One of those visitors returned the following day, did move the sliders, and spent longer on the page, however, this second visit was

excluded in preprocessing step 7, and one must be careful not to cherry-pick the data that confirms one's hypothesis. From this it stands to reason that people do generally explore the larger state-space provided by the extra controllable parameters when they are available, although not universally.

2.4.2. Did Extra Controllable Parameters Increase the Visitors' Curiosity?

About the same proportion of each group, 6 of 10 participants in the *no-sliders* group and 6 of 12 in the *sliders* group, did not click the canvas. There was nothing in the design of the interface that suggested that clicking the canvas would have any effect, nor was doing so necessary to enjoy the piece. Nonetheless, visitors who did so were rewarded with different behavior of the drawing algorithm. I hypothesized that the presence of sliders would make visitors curious to explore whether the canvas was interactive, although this was not the case.

2.4.3. Did Visitors Use the Widget as a Toy, or as a Tool for Making Pictures?

Only 4 participants, two from each group, clicked the "Download" button. This suggests that visitors were generally more interested in the process of interacting with the widget than in the final product of that interaction, i.e., they were using it as a toy and not a tool, as is consistent with the definition of interactive art.

2.4.4. Were the Sliders Engaging?

Participants in the *sliders* group spent more active time on the page ($N = 12$, $M = 75.25$, $SD = 45.45$) than the those in the *no-sliders* group ($N = 10$, $M = 39.90$, $SD = 14.98$). The two-tailed Welch's independent-samples t-test for unequal sample sizes shows that this difference is significant, with $|t(13.77)| = 2.53$, $p < 0.04$. Moreover, this significance is robust in the sense that any sensible variation on the pre-processing steps yields significant results. For example, subtracting out periods of inactivity greater than 5 instead of 10 s, or excluding preprocessing Step 5, both yield $p < 0.04$. This demonstrates that the sliders caused people to engage for longer.

2.4.5. Were Engaged Visitors More Likely to Return?

Preprocessing step 7 might not strictly be the correct approach, as one might hypothesize that an engaging interface would encourage people return more frequently. In fact, when we exclude step 7 from preprocessing, we see that the *sliders* condition had 1.50 visits per participant, while the *no-sliders* condition had only 1.20 visits per participant. Moreover, the difference in the amount of active page time between the *sliders* ($N = 18$, $M = 76.17$, $SD = 41.62$) and *no-sliders* ($N = 12$, $M = 39.08$, $SD = 13.75$) conditions is even more significant, $|t(22.11)| = 3.50$, $p < 0.005$, when including multiple visits per participant. This suggests that not only were *sliders* more likely to return, but when they did return they spent longer than the average on their return visits, while "non-sliders" were less likely to return and spent less time than the average on their return visits. However, the sample size of repeat visitors is small, and thus the observations in the previous sentence are not significant on their own. It could just as well be that a few people who are intrinsically predisposed to visit frequently and spend longer time were assigned to the *sliders* condition by chance.

2.5. Discussion

These results show that providing extra controllable parameters does make interactive art more engaging. However, it is not clear what the limit is; certainly visitors could not be engaged for any arbitrarily long period of time simply by supplying an appropriately large number of controllable parameters. Moreover, one may note that the *no sliders* condition effectively had 0 controllable parameters for visitors who did not click the canvas. Further research is needed to determine the curve that relates engagement to controllable parameters.

3. STUDY 2—FANTASY

Some but not all interactive artworks incorporate fantasy. Malone (1982) defined fantasy in this context as the showing or evoking of "images of physical objects or social situations not actually present". I will adopt the somewhat broader definition that fantasy is the evoking of *anything* that is not actually present. Malone showed that fantasy is a powerful tool for engagement in educational computer games, with the caveat that the fantasy must appeal to the particular visitor. In the domain of interactive art, many responsive environments make clear use of fantasy. In *Connected Worlds* at The New York Hall of Science (Mallavarapu et al., 2019), virtual "water" is projected onto the floor, and visitors can change how it flows by placing real physical obstacles in its path. The fantasy is that there is real water flowing. In *Born From the Darkness a Loving, and Beautiful World* (Sisyu + teamLab, 2018), the fantasies are more abstract. Visitors can interact with projected animations of text, flowers, butterflies, and lightning as if they were tangible. The fantasy is that these objects are tangible. Other responsive environments do not make use of fantasy. In *Fibres Out of Line* (Krzyzaniak et al., 2021), visitors can make a room full of robots play music by moving around in front of a camera. Although some of the robots are fanciful in appearance, the visitors are not meant to imagine anything beyond what is physically present. This raises the research question for Study 2:

Does the presence of fantasy make interactive art more engaging?

3.1. Design

In a previous paper, I describe a words-to-music synthesizer that I designed (Krzyzaniak, 2020), and it occurred to me that it could be repurposed to test fantasy in the context of interactive art. The interface to the synthesizer is depicted in **Figure 2**. There is a text-input field that initially reads "Enter Some Descriptive Text," and there is a graph that shows some default words plotted according to their valance & arousal (sentiment). Visitors can enter words into the text input field, and the software computes and plots the emotional valence and arousal of each word individually, replacing the default words, as well as an average valence and

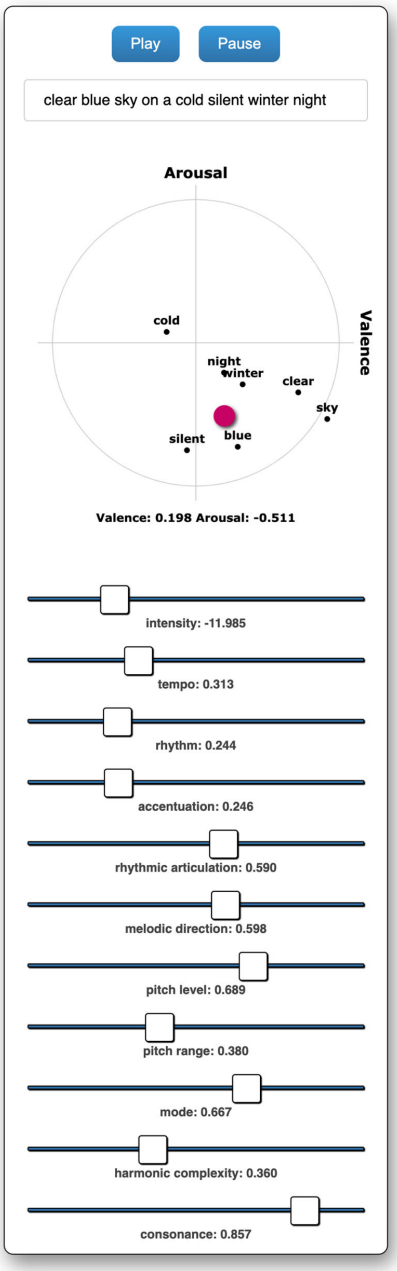


FIGURE 2 | The widget used in Study 2.

arousal score for all of the words taken together (the pink dot). At any point, the visitor can press the *Play* button, and the software will synthesize music in real time that ostensibly matches the average valence and arousal score of the text. Additionally, visitors can directly adjust the musical features using a bank of sliders, or they can manually set the valence and arousal of the music by dragging the pink dot around within the valence & arousal plot, which in turn moves the sliders to some empirically determined values. In order to test the effect of fantasy on engagement, I used this interface as the basis of a new study with two conditions, which I will call *words* and *no-words*. In the *words* condition, visitors were presented exactly the interface shown in **Figure 2**. The *no-words* condition was identical, except that the text input field at the top was not present, and no words were ever plotted in the valence/arousal widget. The theory is that the presence of the text input box encourages visitors to supply their own fantasy, to imagine scenarios, settings or events, and enter them in order to hear what the synthesizer will produce for them. Visitors in the *no-words* condition can still produce the same sounds by manipulating the sliders, but the numerical settings of the sliders will not originate in their fantasies. The widget is available on the internet, and the individual conditions can be accessed *via* the following URLs.

1. redacted_for_anon_review
2. redacted_for_anon_review

3.2. Data Collection

I presented the widget as a demo poster, during an online poster session at a virtual conference on digital musical instruments (NIME 2020). Again it was posted to a university webpage. Visitors were assigned randomly to the two conditions and data was collected as before. It is worth pointing out that the words-to-music synthesizer was originally intended as a tool for sound designers who might, for example, enter part of a movie script and generate background music. Consequently it was not designed to be an interactive artwork by itself. However, whether a given system will be received as tool or a toy sometimes depends on who the visitor is, and under what circumstances they are using it. In this study, because of the setting, the attendees were not using the synthesizer as a tool for making background music, they would have been primed to think of it as a musical instrument, and used it as a toy while browsing poster presentations.

3.3. Data Preprocessing

During the trial period, 69 browsers visited the page a total of 84 times, excluding anyone that had at any point been flagged as a developer in the database. To the data I applied the same preprocessing steps as described in section 2.3 above, with a few small modifications.

1. First, In Steps 3 and 4, as long as the synthesizer was playing, the page was considered active even when the page was not in focus, and even in the absence of input events. Playing means

that the visitor had pressed the *Play* button more recently than the *Pause* button.
2. Moreover, I excluded all visits in which the visitor never pressed the *Play* button at all. There was one visitor in the *words* condition who entered the sentence "angry spiky cactus with poisonous spines," but did not press *Play*, who was excluded in this step. Although it is tempting to include this visit, doing so would apply this step asymmetrically to the conditions, as there is no equivalent check for interactivity in the "no words" condition. In any event, the choice to include or exclude this one participant has no effect on the significance levels of any of the results.
3. Finally, Step 6, which excludes visits less than 20 s in duration, was not performed, as excluding visitors that did not press *Play* obviated the need for this.

After preprocessing, there remained a total of 47 visits, 20 of which were assigned randomly to the *words* condition, and 27 to *no-words*.

3.4. Results

3.4.1. Did Visitors Employ Fantasy When They Could?

A sizable minority of visitors in the *words* condition, 8 out of 20, did not enter any words into the text input field. Six of those moreover did not move the pink dot within the valence/arousal plot which had default words printed on it. This shows that although these six participants did engage with the music by pressing *Play*, they did not engage with the fantasy at all. This is perhaps due to the conference setting, where most people visited this widget during the designated poster session; some visitors probably went quickly from poster to poster, giving only a cursory glance to some posters. This group is interesting, and I will present further analysis on this them in the following subsection.

This notwithstanding, the majority of people that were presented the option to make use of words did so. Most people entered adjectives one at a time, for example mysterious, charismatic, romantic, sexy, crazy, talkative, lively, fucked, diatonic, abstract, uninspired, and tragic. Very few people entered complete sentences, such as "What do you like to eat today?" and "I am so tired." Because of the conference setting, I suspect that most visitors in this condition were in a sense testing or probing the software, to see if they agree with what the synthesizer produces for a given word. This involves imagining the sensation invoked by the word so that it can be compared to the sensation evoked by the synthesizer, and consequently, this qualifies as fantasy under the given definition.

3.4.2. Is Fantasy Engaging?

Participants in the *words* group spent more active time on the page ($N = 20$, $M = 160.5$, $SD = 148.8$) than the those in the *no-words* group ($N = 27$, $M = 77.85$, $SD = 77.48$); about twice as long on average. The two-tailed Welch's independent-samples t-test for unequal sample sizes shows that this difference is significant, with $|t(26.61)| = 2.27$, $p < 0.04$. From this it follows that people are engaged by interactive art that encourages them to fantasize. This result comes with one caveat; In the previous subsection I mentioned that eight people who had the option to enter words did not do so. Looking only within the *words* condition, the people who chose to enter words spent much more time on the page ($N = 12$, $M = 222.5$, $SD = 160.26$), three times longer on average, than those who chose not to enter any words ($N = 8$, $M = 67.38$, $SD = 57.01$). The same Welch test shows that this difference is significant, with $|t(14.74)| = 3.07$, $p < 0.01$. In fact, people in the *words* condition who chose not to enter any words spent about the same amount of active time on the page as those in the *no-words* condition. This highlights the point that encouraging people to fantasize is not sufficient, and a person must also choose to participate in the fantasy.

3.4.3. Is Fantasy Distracting?

No. Visitors in both the *words* and *no-words* groups spent, on average, 68% of their active time listening, without even 1 percentage point difference between the groups. Listening is defined as the total amount of time during which the *Play* button had been pressed more recently than the *Pause* button. This demonstrates first that the extra time spent by visitors in the *words* condition was not attributable to them exploring the words in the absence of music. Nor were they so distracted by the words that they in general felt compelled to pause or defer listening to the music so they could focus on the fantasy. From this it stands to reason that the fantasy contributed to their listening and did not distract from it.

3.5. Discussion

These results show that for some visitors, fantasy has no effect, and for others it is a powerful tool for promoting engagement. In the latter case, the fantasy does not distract visitors away from the rest of the work, but rather they incorporate the fantasy into the overall experience. This demonstrates that the additional time spent on the page is not attributable to the mere presence of an additional page element (text input field) but is in fact a result of the fantasy.

4. STUDY 3—TIMESCALES

Some interactive artworks respond on different timescales than others. Some respond only instantaneously to the immediate actions of the interactant. Others by contrast may continue to respond for some time after the interactant performs an input action. Likewise, in some works a interactant may need to perform some action continuously over a period of time before the artwork begins to respond. This is illustrated in several works of the artist Rafael Lozano–Hemmer[5], which are representative of an entire genre surrounding the idea of digital mirrors.[6] Works like *1984x1984* and *Eye Contact* essentially display a digitally-mediated live video stream of the interactant on a screen. At each frame of video, what is displayed on the screen is determined by the interactant's location and pose at that exact moment in time. *Airborne* and *From Selfie to Self Expression*, are similar, but also have fluid dynamics simulation overlain; interactants can perturb the "fluid" with their motions. In this

[5] All of the works discussed here are documented on his website, https://lozano-hemmer.com/videos.php.

[6] Other notable artists in this genre are Daniel Rozen, Golan Levin, and Zach Lieberman.

way, the actions of the interactant's continue to have an effect for some time after they are performed. In *People on People*, an interactant's current silhouette may be superimposed with videos of themselves recorded moments previously, allowing them to interact with past versions of themselves through the work. Thus, the interaction unfolds over a period of time. Other works may respond to the *average* behavior of the interactant. *Particle Falls* by Andrea Polli[7] visualizes air pollution, so that in principle many people would need to change their behavior over a long period of time to have a large effect. The research question for Study 3 is, therefore:

Is there an optimal timescale that engages people the most?

4.1. Design

To test this question in a controlled environment, I developed the widget depicted in **Figure 3**. It consists of a blank canvas and some sliders. When a visitor touches or clicks down on the canvas, the tip of a metaphorical pen begins drawing a colorful spirograph curve, with the pen trace orbiting around the finger or cursor location. If the finger or cursor is dragged within the canvas, the orbital center of the curve follows. A second, mirror-image, grayscale spirograph curve is drawn at an opposing location on the canvas. The curves fade out over time as they are drawn, so that at any moment in time only recently drawn portions of the curves are visible, with progressively older portions of the curves appearing progressively fainter until sufficiently old portions of the curves do not appear at all. When the visitor releases the click or stops touching the canvas, the pen tips continue drawing the curves for some time, but their speed decreases and eventually stops, at which point no new length is added to the curves. If the finger or cursor was being dragged at the time of the release, the orbital centers of the curves continue moving inertially within the canvas for some distance. Additionally, visitors can adjust the sliders, which control some parameters pertaining to how the curves are drawn. Adjusting any slider also has the effect of causing a portion of the spirograph curve to be drawn so that the effects of the parameter can be seen.

There are four conditions. In condition 0, the time it takes for a portion of curve to fade completely out, the time it takes for the pen velocity to go to zero when the click or touch is released, and the time it takes for the orbital centers to come to rest, are all less than 1 s in duration. In condition 1, they are approximately 3 to 5 s in duration. In condition 2 they are approximately 10 to 15 s. In condition 3 they are infinitely long, such that once the visitor touches the canvas or moves a slider, the pens will continue to wander around the canvas forever, eventually filling every pixel, similar to the animation in Study 1. These increasingly long durations represent increasing timescales as described in the introduction to this section. **Figure 3** depicts condition 2.

The widget is available on the internet, and the individual conditions can be accessed *via* the following URLs.

1. redacted_for_anon_review
2. redacted_for_anon_review
3. redacted_for_anon_review
4. redacted_for_anon_review

[7]http://eco-publicart.org/particle-falls/

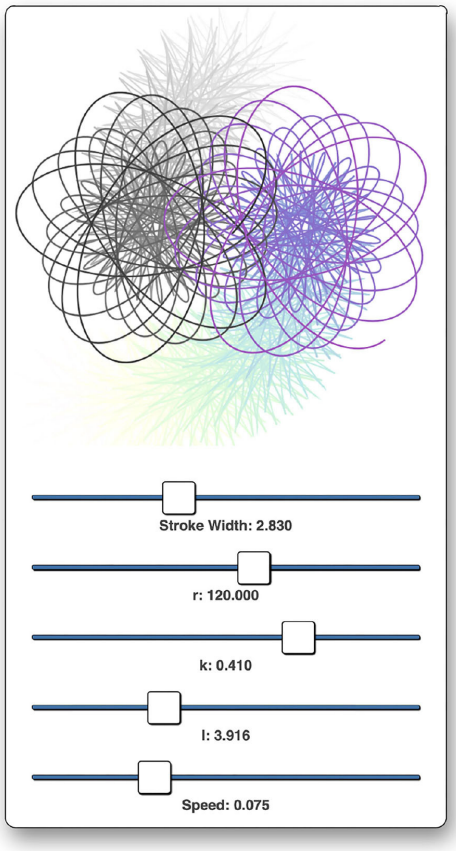


FIGURE 3 | The widget used in Study 3.

4.2. Data Collection

I posted the widget to my university biography page as in Study 1, and made no specific recruitment efforts aside from including the link in the bottom of all of my emails. I left it there with no other page content other than the required page template as discussed above for a period of 10 months beginning in April 2020.

4.3. Data Preprocessing

Over the trial period, 227 browsers not belonging to registered developers visited the page a total of 354 times. To these, I applied the preprocessing steps as described in section 2.3 above, with a few modifications, as follows.

1. First, I only included "interactive" visits. To be considered interactive, the visitor had to either click on the canvas or adjust one of the sliders at least once. Determining if a visit was interactive was performed after joining visits separated by less than 10 s.
2. Additionally, Step 6, which excludes visits less than 20 s in duration, was not performed, as excluding non-interactive visits obviated the need for this.

The majority of visits, 80% of them, were not interactive, with only 75 interactive visits from 66 distinct browsers. Again I only consider the first visit by each browser unless otherwise stated. Thus in total, after preprocessing, there remained 66 visits by those 66 browsers, with 12, 18, 12, and 24 visitors assigned randomly to conditions 0, 1, 2, and 3, respectively. This accounted for a cumulative total of 67 min and 5 s of active time on the page.

4.4. Results

4.4.1. Did People Engage for Longer in the Conditions With the Longer Timescales?

No. On average across all conditions, each visitor spent 61 active seconds on the page with a relatively large standard deviation of 51 s. I hypothesized that longer timescales might stretch out the visitors' attention, causing them to spend longer on the page. However, comparing the conditions pairwise using a two-tailed Welch's independent-samples *t*-test for unequal sample sizes showed that there was no significant difference between conditions. Nowhere was p even as small as 0.5, nor the confidence as great as 50%, so the results of these comparisons were exceptionally insignificant. From this it follows that the longer timescales had no effect on how much time people spent on the page, and it is not likely that any minor variation on this study would yield significant results.

4.4.2. Did People Click the Canvas More?

People clicked the canvas more in conditions 0 ($N = 12$, $M = 5.4$, $SD = 5.2$) and 3 ($N = 24$, $M = 6.5$, $SD = 13.5$) than in conditions 1 ($N = 18$, $M = 1.8$, $SD = 3.5$) and 2 ($N = 12$, $M = 2.8$, $SD = 4.9$). This appears to result in a U-shaped curve representing number of clicks as a function of the timescale. This could indicate that the timescale affects the *style* of interaction. For intermediate timescales, visitors perform periodic actions and then pause to observe the effects, whereas for extreme timescales, visitors continually perform actions to try to keep exerting influence over the system. By contrast, visitors on average made a total of 5 or 6 slider adjustments regardless of condition (adjusting each of the 5 sliders approximately once). A slider adjustment means that they moved and released the slider. This shows that the timescales did not influence the visitors's overall curiosity to explore the piece despite the ostensibly different styles of interaction represented by different clicking patterns. However, the two-tailed Welch's independent-samples *t*-test for unequal sample sizes shows that the differences in the number of clicks per condition are only marginally significant, with conditions 0 and 3 taken together and compared against conditions 1 and 2 yielding $|t(45.28)| = 1.91$, $p < 0.1$. Further research with a larger sample size is needed to clarify whether this effect is real.

4.5. Discussion

The examples in the introduction to this chapter should make it clear that "timescales" refers to a variety of different but related concepts. This study primarily tested the concept of perturbing a system such that actions continue to have effect into the future. Overall this has no effect on engagement for the timescales studied, but might affect how people interact with the work. The other similar concepts could be tested separately in the future.

5. STUDY 4—AGENCY

Many interactive artworks have some sort of *agency*. Throughout this section, I will refer to an artwork that ostensibly has agency as an "agent." Agency is defined here to be the ability for an agent to act upon the world (Russell and Norvig, 2002).[8] Moreover, these actions must be done deliberately, in order to accomplish something; and spontaneously, without external stimulus (Wooldridge and Jennings, 1994). Insofar as agency is a property of the agent, it may manifest itself in a few different ways. In interactive art, agency often means that the agent has some behaviors that are only partially influenced, but not fully controlled, by the interactant's actions (Dahlstedt, 2021); for example an interactive musical robot that sometimes mimics musical themes that it heard, but other times introduces novel and appropriate material not related to what it heard. The new material was produced spontaneously, and, if it is not completely random, deliberately. Agency may also manifest itself as the use of action to express a (perceived) mental state, such as emotion or desire (Misselhorn, 2015), for example a robot that smiles at people wearing hats and frowns at everyone else. The actions of smiling and frowning are deliberate in the sense that it accomplishes something (expressing like of hatted people). Even though these actions are in response to a person's presence, they are nonetheless spontaneous in the sense that they are driven by the robot's own inner state. Furthermore, agency is also a property of the interactant, because whatever the agent's properties, the interactant must have a certain theory of mind with regard to the agent, otherwise its actions will appear random and meaningless, instead of deliberate, directed, and purposeful. Ultimately an agent only has agency if the interactant ascribes agency to it (Takayama, 2012).

These principles are illustrated by two works of Golan Levin.[9] *Opto-Isolator* is a robotic eye that follows you as you move around, and blinks whenever you blink. It has little or no agency as it does not appear to initiate action or have any behaviors that are not fully controlled by the interactant's actions.[10] *Snout* is another robotic eye, but it is different in that it appears to look around, only sometimes focusing on the interactant and

[8]Not that "agency" more typically refers to the *interactant's* ability to act within the system; this is a separate question not considered here.

[9]The works here can be seen in his Ted Talk, https://www.ted.com/talks/golan_levin_art_that_looks_back_at_you.

[10]The artist says that it may look away if you look at it for too long, which may imbue it with a small amount of agency.

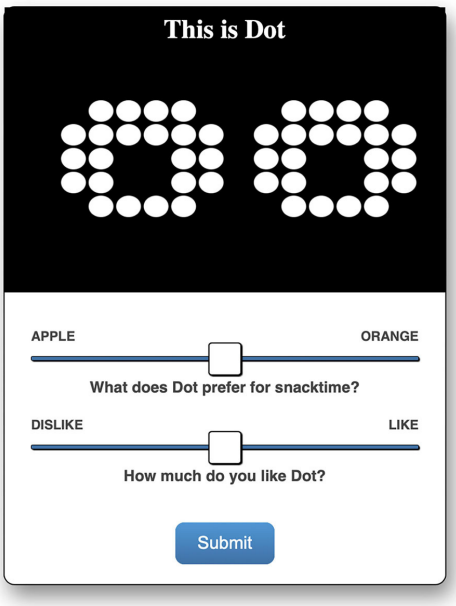


FIGURE 4 | The widget used in Study 4.

sometimes not. This is because it is outdoors and is sometimes distracted by trees or other movement in the environment. This gives it the impression of having some internal process that is only partly influenced by the interactant's. Additionally, it sometimes recoils in a surprised gesture, which is an action that expresses an internal state. Due to these features, I personally ascribe greater agency to *Snout* than *Opto-Isolator*. This gives rise to the research question for Study 4:

Does ascribed agency promote engagement in interactive art?

5.1. Design

To test the hypothesis that greater ascribed agency leads to greater engagement, I designed the widget shown in **Figure 4**. The widget shows a representation of two low-resolution LED robot eyes, similar to the eyes of robots such as Eve in Pixar's WALL-E, the toy robot Cozmo by Anki, and, most saliently, my Dr. Squiggles robot (Krzyżaniak, 2021). Above the eyes is the statement "This is Dot." Beneath the eyes is a survey form consisting of two questions and corresponding sliders, implicitly ranging from 0 on the left to 1 on the right, and a submit button. The first question, which I will henceforth call the *agency* question, asks whether Dot prefers apples or oranges for snack time. The second question, which I will call the *likability* question, asks how much you like Dot. When visitors press the submit button, a message is displayed that either thanks them, or prompts them to move both sliders before submitting, if they have not yet done so. This study has four conditions. In the *control* condition, the eyes are presented as a static image that do not move, exactly as depicted in **Figure 4**. In the second condition, the *two eye* condition, the eyes are animated. They track the position of the cursor as the visitor moves it around the page, and in particular they appear to watch the visitor as they adjust the sliders. I accomplish this by offsetting both the location of the pupil within the eye, and the location of the eye within the widget, in the direction of the cursor by an amount proportional to the distance from the cursor to the center of the widget. Moreover, in this condition, immediately after the visitor moves and releases the *agency* slider, the eyes attempt to indicate a preference for the position of the slider. If the slider is placed in the left half of the range, the eyes move rapidly back and forth to indicate "no." If the slider is placed in the right half of the range, the eyes move from rapidly from left to right several times indicating that the slider should be moved even further right, unless the slider is placed in the rightmost 10% of the range, in which case the eyes move up and down to indicate "yes." There is a third *one eye* condition in which there is only one eye, and the size and shape are nearly identical to the design used in Dr. Squiggles. This eye has the same behavior as in the *two eye* condition. The fourth and final *angular offset* condition is identical to the *two eye* condition, except that instead of the position of the pupil and eyes being offset directly in the direction of the cursor, they are offset in the direction of the cursor plus some angle. The measure of the angle drifts over time using Brownian motion, unless the cursor is in the vicinity of the sliders, in which case the angle is zero so the eyes appear to be watching the visitor adjust them. I will refer to the three non-control conditions collectively as the *animated* conditions. For reference, the widget is available on the internet, and the various conditions can be visited using the following URLs:

1. redacted_for_anon_review
2. redacted_for_anon_review
3. redacted_for_anon_review
4. redacted_for_anon_review

5.2. Data Collection

This study is somewhat different from the others in that it is clear by looking at it that it is a study, which made it easier to recruit participants. I uploaded the widget to my personal website. Because my personal website does not force visitors to "consent" to site-wide data-collection, I included a small link at the bottom of the page explaining the study. I emailed a link to the widget to a large professional mailing list, asking participants to participate in a 2-question study. I let the study collect data for about a week.

5.3. Data Preprocessing

During the trial period, 122 browsers visited the page a total of 143 times, excluding anyone that had at any point been flagged as a developer in the database. I measured the active time the visitors spent on the page using the same preprocessing steps as described in section 2.3 above, with a few small modifications, as follows.

1. First, because the eyes follow the cursor, I removed all visits by touchscreen devices for which this would not work as intended. A device was considered to be a touchscreen device if the touchstart, touchend, or touchmove Javascript user interface events fired anywhere on the page prior to any mousedown, mouseup, or mousemove events. This resulted in the removal of 16 devices.
2. Visits were only included if they were submitted. Submitted means that the submit button had been pressed after adjusting each of the sliders.
3. Additionally, Step 6, excluding visits of less than 20 active seconds, was not performed, as it is plausible that some valid visitors would have spent less than 20 s completing the survey. Excluding non-submitted responses obviated the need for this step.

In addition to collecting the active page time, I recorded each adjustment of each slider and each press of the submit button, irrespective of the order of those events. To be clear, pressing the submit button did not actually submit the responses, it only recorded the fact that the visitor had pressed it, and all data were committed once the visitor closed or navigated away from the page, so the active page time could be captured. After preprocessing, there were 85 responses from 85 visitors, with 18, 24, 22, and 21 participants assigned to the *control*, *two eyes*, *one eye*, and *angular offset* conditions, respectively.

5.4. Results

5.4.1. Did Visitors Notice That Dot Responded to the *Agency* Slider?

Only some did. In the three *animated* conditions taken together, visitors on average moved the *agency* slider a greater number of times ($N = 67$, $M = 3.99$, $SD = 5.02$) than in the *control* condition ($N = 18$, $M = 1.67$, $SD = 0.91$). The two-tailed Welch's independent-samples t-test for unequal sample sizes shows that this difference is significant with $|t(78.52)| = 3.57$, $p < 0.001$. The same is also true for the *likability* slider, with ($N = 67$, $M = 2.43$, $SD = 3.91$) and ($N = 18$, $M = 1.22$, $SD = 0.43$), respectively, and $|t(71.46)| = 2.48$, $p < 0.02$. These facts suggest that the animation made people curious to explore both sliders. Moreover, within the three *animated* conditions taken together, the same Welch's test shows that the average number of times that visitors moved the *agency* slider was significantly higher than the number of times they moved the *likability* slider, with $|t(124.5)| = 2.00$, $p < 0.05$. So although they engaged more with both sliders in the *animated* conditions, they did so disproportionately more with the *agency* slider. This suggests that the visitors did on average notice that Dot responded to the movement of that slider and not the *likability* slider. They played with it to further explore the interaction.

Having said that, about 50% of visitors in all conditions together, and in each one separately, moved the *agency* slider only once, which was required in order to successfully press the submit button. They did not subsequently make many adjustments to it in response to Dot's actions. An initial pilot of this study amongst colleagues suggested that many visitors with this profile in the animated conditions did not notice that Dot responded to the *agency* slider. So although the average visitor did notice, only half of individual visitors did. In this study, noticing this action was a prerequisite for the ascription of agency, since Dot used this action to indicate that it *wants* something (an orange and not an apple). Visitors who did not notice the interaction could not have possibly ascribed agency to Dot. This is somewhat different than noticing the action but not believing it to be purposeful.

5.4.2. Did Visitors Ascribe Agency to the Movement Associated With the *Agency* Slider?

Here I will operationalize the amount of ascribed agency as the final position of the *agency* slider at the time visitors navigated away from the page. The slider will on average be biased to the right *iff* (a) Dot acts in such a way as to express a rightward preference for the slider position, and (b) visitors ascribe desire to these actions, as opposed to interpreting them as arbitrary.

Looking only at visitors who moved the *agency* slider more than once, in the *angular offset* condition the average position of the *agency* slider at the time visitors navigated away from the page was further to the right ($N = 10$, $M = 0.86$, $SD = 0.31$) than in the *control* condition ($N = 8$, $M = 0.44$, $SD = 0.41$). The two-tailed Welch's independent-samples t-test for unequal sample sizes shows that this difference is significant with $|t(12.82)| = 2.44$, $p < 0.04$. The same was not true for the *likability* slider which had a final position of about 0.69 in both conditions. This suggests that these visitors understood that Dot wanted them to move the *agency* slider but not the *likability* slider to the right. Understanding that an agent wants something is equivalent to ascribing agency to it under the given definition.

Again looking only at visitors who moved the *agency* slider more than once, in the *two eye* and *one eye* conditions, the final value of the *agency* slider was similarly higher than in the *control* condition with ($N = 14$, $M = 0.77$, $SD = 0.27$) for the *two eye* and ($N = 12$, $M = 0.66$, $SD = 0.37$) for the *one eye* condition. However, these differences were not significant. Using a weaker test, 12 out of 14 participants in the *two eye* condition left the *agency* slider in the right half of its range; the probability of at least this many people doing so by chance alone is less than 1%, as compared to exactly half of visitors in the *control* condition doing this. This suggests that visitors in the *two eye* condition in general did ascribe agency, although more weakly, as they only partially understood or complied with Dot's desire that they move the slider all the way to the right. In other words, these visitors likely interpreted some of Dot's actions as random and not deliberate. In the *one eye* condition, 8 out of 12 visitors left the *agency* slider in the right half of its range, which would occur with 19% probability by chance alone. This suggests that although these visitors did on average notice that Dot responded to them moving the *agency* slider, many did not understand that Dot was asking them to do something, meaning that they ascribed little or no agency to Dot. For completeness, 9 out of 10 participants in the *angular offset* condition did this, with about 1% chance of happening by

accident, confirming again that the visitors ascribed agency in this condition.

It is difficult to compare between the *animated* conditions, because the differences are slight. However, these findings may suggest that visitors ascribed the most agency in the *angular offset* condition, followed by the *two eye* condition, then the *one eye* condition. The angular *offset condition* might be explained by the fact that it was the only condition in which Dot had some continual process that was only partially affected by the visitors's actions. The continual interplay between the visitor and Dot may have primed visitors to think of Dot as an agent. By contrast, visitors in the *two eye* condition clearly understood that Dot was asking them to move the slider to the right, but were not as attentive to all of the signals it was giving about how far to the right they should move it. Nonetheless, the *two eye* condition is slightly more anthropomorphic than the *one eye* condition, which might explain why so little agency, if any, was ascribed in that condition.

5.4.3. Did Any Visitors Deliberately Oppose the Dot's Desire?

No. Of the 36 visitors in the three *animated* conditions who moved the *agency* slider more than once, only one visitor did leave it to the extreme left of its range below 0.05 at the time of navigating away from the page, and they moved it there after the last time they pressed submit. By contrast, 16 of these visitors did leave it to the extreme right above 0.95. This suggests that in general people did not antagonize Dot. By contrast, out of the 31 visitors in those three conditions who only moved the *agency* slider once, 8 did leave it to the extreme left and 6 to the extreme right. This is expected since the first placement of that slider is random.

5.4.4. Did Visitors Prefer Two Eyes Over One?

In addition to the one-eyed artworks discussed in the introduction, the authors of this paper have independently developed one-eyed musical agents (Erdem, 2021; Krzyżaniak, 2021). Although it is somewhat tangential, we wanted to know if people expresses a greater preference for two-eyed agents. This appears not to be the case, with the average position of the *likability* slider at the time visitors navigated away from the page being 0.66 in all conditions combined, with no significant differences between conditions.

5.4.5. Did Visitors Engage for Longer When They Ascribed Greater Agency to the Eyes?

Yes. In the three *animated* conditions taken together, visitors spent more active time on the page ($N = 67$, $M = 48.50$, $SD = 22.46$) than in the *control* condition ($N = 18$, $M = 30.11$, $SD = 14.68$). The two-tailed Welch's independent-samples t-test for unequal sample sizes shows that this difference is significant with $|t(40.93)| = 4.16$, $p < 0.001$. The same is true for each *animated* condition taken separately and compared to the non-animated condition, with $p < 0.01$ in each case, and no significant difference between the *animated* conditions. But did people spend longer in these conditions only because they were interactive, or specifically because that interaction involved agency?

Considering all 67 visitors in the three *animated* conditions, there was a weak but significant positive correlation between the final position of the *agency* slider and the amount of active time spent on the page, with $r(65) = 0.32$, $p < 0.01$. By contrast, there was no correlation between the like *likability* slider and the active page time, with $r(65) = 0.19$, $p > 0.1$, and if anything the trend was slightly negative. Similarly in the *control* condition, the final position of neither the *likability* nor *agency* slider had a significant correlation with page time, with both having a slightly negative trend. From this it follows that greater ascribed agency was associated with more engagement. The equation for the relationship is y=18.46x + 37.13, where y is page time in seconds and x is the final *agency* slider position, from 0 on the left to 1 on the right. This means that visitors in the *animated* conditions who ascribed no agency because they did not even notice Dot's actions spent on average 37 s on the page, as compared to the 30 s average in the *control* condition. The extra 7 s are attributable to the interactivity alone, with an additional 18 s spent by visitors who ascribed the most agency to that interactivity. From this it stands to reason that for the average visitor, agency is about as powerful at promoting engagement as simple interactivity, and the two are additive. Note however that it is not known whether people spent longer because of the agency, or instead if people who stayed longer for other reasons ended up ascribing more agency.

5.5. Discussion

In this section, we have observed that about half of people failed to notice, in a fundamental way, what was going on in the study. This mirrors the finding in Study 2 regarding fantasy, that presenting visitors with the *opportunity* to fantasize or ascribe agency isn't sufficient; visitors must also be receptive and willing to engage in that way. Of those who did notice, some ascribed more agency than others, and this may be due to anthropomorphism, and to the presence of some behaviors that are only partially controlled by the interactant, although these are both subtle and probably very complex, and likely a great amount of additional research will be needed to tease this apart convincingly. Whatever the reason, visitors who ascribed the most agency also engaged for the longest. Finally, agency may be useful for directing people's behavior, since people who noticed what was going on in the study generally complied with Dot's desire, and did not antagonize Dot. This shows that agency can be a powerful tool for completing the feedback loop between the interactant and the work.

6. CONCLUSION

To briefly recapitulate, the studies herein have shown that (a) more controllable parameters increase engagement; (b) fantasy strongly increases engagement for some people but not at all for others; (c) timescales do not influence engagement but might affect the style of interaction, and (d) ascribed agency is related to increased engagement. Note, however, that this should not be taken as a comprehensive framework for how to promote

engagement in interactive art. These are only a small sampling of what is undoubtedly a myriad of properties that might promote engagement. Even the few properties presented here are very complex and the studies in some sense raise more questions than they answer. Therefore this paper should be taken as a starting point, not an end point.

This paper has left open many avenues for future work, beyond extending similar methods to other properties of art. The limited data collected in the studies is both a strength and a weakness of the presented method. On the one hand it has allowed us to carefully control the experiments in an ecologically valid setting. On the other hand, we are viewing the visitors through a pinhole, and there is a lot that we just don't know. All art is inherently cultural, and experiencing it depends on enculturation, but we do not know the demographics of the participants in the studies because we did not collect that information. We don't know why some people stop to interact with a widget when presented with it, and others just leave the page without engaging at all. We don't know whether people engaged socially, for instance if two people interacted with a widget together on the same web browser. We don't know what metacognitive processes people may have engaged in during interaction. We don't know what role memory and learning may have had in the interactions, as would be especially applicable to repeat visitors. We do not know the longer-term effects of the interactions, for example if an interaction caused a shift in perspective that altered a participant's behavior in their daily life at a later date. All of these are avenues for future work, both because they are interesting questions in their own right, and because some extra information would improve the repeatability and accuracy of studies of this nature.

As a final note, it is interesting to think about how these results would apply to other types of systems, especially more complex ones. The authors have a special interest in interactive *musical* systems like musical robots, responsive dance works, and musical software agents. Even knowing that fantasy is important, it is not clear, for example, how the design of a guitar robot's body might encourage or discourage fantasy in its musical partners. When a robot improvises music with a human partner, what is the optimal level of ascribed agency so that its playing is neither to predictable nor too random, and how can that be achieved? How can these and other properties be combined in a system that is enjoyable to play music with, that helps people learn an instrument, or that otherwise helps people reap the benefits of lifelong music making?

Taking a step back, interactive art in general clearly has great potential for engagement. The average 27 s people spent looking at paintings (and reading the label) in Smith and Smith (2001) included some of the greatest masterpieces in history, and people reported having transformative experiences while looking at them. By contrast, none of the groups reported in this paper spent a mean of less than 30 s interacting with the artwork, even in the control conditions. In fact, double that time was common, with about a minute seeming like the default. One group even spent 222 s on average—more than 8 times as long as people spend looking at paintings; and these are not masterpieces by any stretch. This demonstrates that interactivity itself is a powerful tool for engagement. However, the great variability across the groups in this article highlights that engagement does not come for free in interactive art. The art must also be thoughtfully designed to have the right properties, including but certainly not limited to the ones presented in this paper, in order to promote engagement.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MK was solely responsible for the first three studies. ÇE and MK contributed equally to the fourth studies, with key insights coming from ÇE. KG provided supervision and feedback. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

REFERENCES

Blythe, M., and Monk, A. (2018). *Funology 2 From Usability to Enjoyment, 2nd Edn*. Cham: Springer. Available online at: https://link.springer.com/book/10.1007/978-3-319-68213-6

Bongers, B., and Mery, A. (2011). "Interactive kaleidoscope: audience participation study," in *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (Canberra, ACT), 58–61.

Chou, C., Condron, L., and Belland, J. C. (2005). A review of the research on internet addiction. *Educ. Psychol. Rev.* 17, 363–388. doi: 10.1007/S10648-005-8138-1

Costello, B., and Edmonds, E. (2007). "A study in play, pleasure and interaction design," in *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces* (Helsinki), 76–91.

Costello, B. M., and Edmonds, E. A. (2009). "Directed and emergent play," in *Proceedings of the Seventh ACM Conference on Creativity and Cognition* (New York, NY), 107–116.

Dahlstedt, P. (2021). "Musicking with algorithms: thoughts on artificial intelligence, creativity, and agency," in *Handbook of Artificial Intelligence for Music* (Cham: Springer), 873–914.

Desmet, P. (2003). "Measuring emotion: development and application of an instrument to measure emotional responses to products," in *Funology* (Dordrecht: Springer), 111–123.

Erdem, Ç. (2021). *First a Guitarist, Then a Drummer Plays with CAVI*. Available online at: https://www.youtube.com/watch?v=WuZBXUpn60Q (accessed October 03, 2021).

Fernaeus, Y., Höök, K., and Ståhl, A. (2018). "Designing for joyful movement," in *Funology 2* (Cham: Springer), 193–207.

Ishii, H., Ratti, C., Piper, B., Wang, Y., Biderman, A., and Ben-Joseph, E. (2004). Bringing clay and sand into digital design - continuous tangible user interfaces. *BT Technol. J.* 22, 287–299. doi: 10.1023/B:BTTJ.0000047607.16164.16

Jordà, S., Kaltenbrunner, M., Geiger, G., and Bencina, R. (2005). "The reactable," in *ICMC* (Barcelona: Citeseer).

Karat, C.-M., Karat, J., Vergo, J., Pinhanez, C., Riecken, D., and Cofino, T. (2002). That's entertainment! designing streaming, multimedia web experiences. *Int. J. Hum. Comput. Interact.* 14, 369–384. doi: 10.1080/10447318.2002.9669125

Krzyżaniak, M. (2021). Musical robot swarms, timing, and equilibria. *J. New Music Res.* 50, 279–297. doi: 10.1080/09298215.2021.1910313

Krzyzaniak, M., Gerry, J., Kwak, D., Erdem, C., Lan, Q., Glette, K., et al. (2021). *Fibres Out of Line*. Available online at: https://michaelkrzyzaniak.com/Fibers_Out_Of_Line/ (accessed July 21, 2021).

Krzyzaniak, M. J. (2020). "Words to music synthesis," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Birmingham: Birmingham City University), 29–34.

Ljungblad, S., Skog, T., and Holmquist, L. E. (2003). "From usable to enjoyable information displays," in *Funology* (Cham: Springer), 213–221.

Mallavarapu, A., Lyons, L., Uzzo, S., Thompson, W., Levy-Cohen, R., and Slattery, B. (2019). "Connect-to-connected worlds: piloting a mobile, data-driven reflection tool for an open-ended simulation at a museum," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow), 1–14.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cogn. Sci.* 5, 333–369.

Malone, T. W. (1982). "Heuristics for designing enjoyable user interfaces: Lessons from computer games," in *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (New York, NY), 63–68.

Misselhorn, C. (2015). "Collective agency and cooperation in natural and artificial systems," in *Collective Agency and Cooperation in Natural and Artificial Systems* (Cham; Heidelberg; New York, NY; Dordrecht; London: Springer), 3–24. Available online at: https://link.springer.com/content/pdf/bfm%3A978-3-319-15515-9%2F1.pdf

Monk, A., Hassenzahl, M., Blythe, M., and Reed, D. (2002). "Funology: designing enjoyment," in *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (New York, NY), 924–925.

Overbeeke, K., Djajadiningrat, T., Hummels, C., Wensveen, S., and Prens, J. (2003). "Let's make things engaging," in *Funology* (Cham: Springer), 7–17.

Pagulayan, R. J., Steury, K. R., Fulton, B., and Romero, R. L. (2003). "Designing for fun: user-testing case studies," in *Funology* (Cham: Springer), 137–150.

Rosson, M. B., and Carroll, J. M. (2018). "Fun for all: promoting engagement and participation in community programming projects," in *Funology 2* (Cham: Springer), 507–518.

Russell, S., and Norvig, P. (2002). *Artificial intelligence: a Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Sapolsky, R. M. (2017). *Behave: The Biology of Humans at Our Best and Worst*. New York, NY: Penguin.

Sisyu + teamLab (2018). *Born From the Darkness a Loving, and Beautiful World*. https://www.teamlab.art/jp/w/whatloving-dark/ (accessed June 21, 2021).

Smith, J. K., and Smith, L. F. (2001). Spending time on art. *Empir. Stud. Arts* 19, 229–236. doi: 10.2190/5MQM-59JH-X21R-JN5J

Smith, L. F., Smith, J. K., and Tinio, P. P. (2017). Time spent viewing art and reading labels. *Psychol. Aesthet. Creativity Arts* 11, 77. doi: 10.1037/aca0000049

Sykes, J., and Wiseman, R. (2003). "Deconstructing ghosts," in *Funology* (Cham: Springer), 243–248.

Takayama, L. (2012). "Perspectives on agency interacting with and through personal robots," in *Human-Computer Interaction: the Agency Perspective* (Berlin: Springer), 195–214. Available online at: https://link.springer.com/content/pdf/10.1007/978-3-642-25691-2.pdf

Wooldridge, M., and Jennings, N. R. (1994). "Agent theories, architectures, and languages: a survey," in *International Workshop on Agent Theories, Architectures, and Languages* (Berlin: Springer), 1–39.

Zytko, D., Grandhi, S., and Jones, Q. (2018). "The (un) enjoyable user experience of online dating systems," in *Funology 2* (Cham: Springer), 61–75.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendices

Appendix A

Supplementary Material

A.1 Paper I

- Code: https://doi.org/10.5281/zenodo.6478035
- Video: https://youtu.be/hpECGAkaBp0

A.2 Paper III

- Code: https://doi.org/10.5281/zenodo.6478037
- Video 1: https://youtu.be/_--dzA5pl9k
- Video 2: https://youtu.be/ikan7NbPTAM

A.3 Paper IV

- Code: https://doi.org/10.5281/zenodo.6478033
- Data: https://doi.org/10.5281/zenodo.6470236
- Video: https://youtu.be/-_wgBZY2iF8

A.4 Paper V

- Code & Questionnaires: https://doi.org/10.5281/zenodo.6478027
- Data: https://doi.org/10.5281/zenodo.6470236
- Video: https://youtu.be/WuZBXUpn60Q