Investigating the cognitive foundations of collaborative musical free improvisation: Experimental case studies using a novel application of the subsumption architecture

Thesis

Doctoral Thesis

Investigating the cognitive foundations of collaborative musical free improvisation

Experimental case studies using a novel application of the subsumption architecture

Author:

Adam Linson

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Centre for Research in Computing,
Faculty of Mathematics, Computing and Technology,
The Open University, UK

Supervisors:

Robin Laney (Open University, UK)
Chris Dobbyn (Open University, UK)
George Lewis (Columbia University, USA)

Examiners:

Geraint Wiggins (Queen Mary, University of London, UK)
Neil Smith (Open University, UK)

Submitted 31-Oct-2013
Examined 20-Dec-2013
Revised 24-Feb-2014

Declaration of Authorship

I, Adam Linson, declare that this thesis and the work presented in it are my own.

Abstract

This thesis investigates the cognitive foundations of collaborative musical free improvisation. To explore the cognitive underpinnings of the collaborative process, a series of experimental case studies was undertaken in which expert improvisors performed with an artificial agent. The research connects ecological musicology and subsumption robotics, and builds upon insights from empirical psychology pertaining to the attribution of intentionality. A distinguishing characteristic of free improvisation is that no over-arching framework of formal musical conventions defines it, and it cannot be positively identified by sound alone, which poses difficulties for traditional musicology. Current musicological research has begun to focus on the social dimension of music, including improvisation. Ecological psychology, which focuses on the relation of cognition to agent–environment dynamics using the notion of affordances, has been shown to be a promising approach to understanding musical improvisation. This ecological approach to musicology makes it possible to address the subjective and social aspects of improvised music, as opposed to the common treatment of music as objective and neutral. The subjective dimension of musical listening has been highlighted in music cognition studies of cue abstraction, whereby listeners perceive emergent structures while listening to certain forms of music when no structures are identified in advance. These considerations informed the design of the artificial agent, *Odessa*, used for this study. In contrast to traditional artificial intelligence (AI), which tends to view the world as objective and neutral, behaviour-based robotics historically developed around ideas similar to those of ecological psychology, focused on agent–environment dynamics and the ability to deal with potentially rapidly changing environments. Behaviour-based systems that are designed using the subsumption architecture are robust and flexible in virtue of their modular, decentralised design comprised of simple interactions between simple mechanisms. The competence of such agents is demonstrated on the basis of their interaction with the environment and ability to cope with unknown and dynamic conditions, which suggests the concept of improvisation. This thesis documents a parsimonious subsumption design for an agent that performs musical free improvisation with human co-performers, as well as the experimental studies conducted with this agent. The empirical component examines the human experience of collaborating with the agent and, more generally, the cognitive psychology of collaborative improvisation. The design was ultimately successful, and yielded insights about cognition in collaborative improvisation, in particular, concerning the central relationship between perceived intentionality and affordances. As a novel application of the subsumption architecture, this research contributes to AI/robotics and to research on interactive improvisation systems. It also contributes to music psychology and cognition, as well as improvisation studies, through its empirical grounding of an ecological model of musical interaction.

Acknowledgements

Thank you to my supervisors, Robin Laney, Chris Dobbyn, and George Lewis, for your time, involvement, intellectual engagement, and invaluable guidance and feedback throughout the process of completing this thesis, and for your friendship outside the halls of the university. Just to name a few specifics among many: I am grateful to Robin, for welcoming me into your research group; to Chris, for conversations about both the substance and style of my writing; and to George, for your encouragement and support over nearly twenty years.

Many thanks to the friends who also kindly took on the role of participants in my study: Evan Parker, John Russell, Okkyung Lee, Pat Thomas, Paul Cram, Peter Evans, Sara Schoenbeck, and Ute Wasserman. Thank you for taking interest in the project and for generously offering your time, music, minds, and words (and homes!). Additional thanks to Aleks Kolkowski, Axel Dörner, Jon Rose, Nate Wooley, Philipp Wachsmann, Rudi Mahall, and special thanks to Alexander von Schlippenbach.

I thank Geraint Wiggins and Neil Smith for engaging with my work in the viva exam, and Simon Holland and Robert Samuels for doing so as well in the preparatory exam.

Thanks to Alistair Zaldua and Tom Mudd for helping out with the studio at Goldsmiths, and to Emmanuel Spinelli for engineering the recordings there with the Disklavier. Thanks are also due to the Vortex Jazz Club in London, including George Coote, Jon Russell Brown, and Oliver Weindling.

Thank you to Martyn Hammersley for assistance in matters of methodology; to Rebecca Fiebrink and the Princeton Computer Science department for the opportunity to present my work; and to Allan Jones for inviting me to present to the OU's Society and Information Research Group, whose members are also thanked for their lively discussion. I also wish to thank Anselm Berrigan for helping me to transform my understanding of the writing process.

Thanks to family, friends, and colleagues for encouragement along the way, with much appreciation to Caroline Forbes, Joel Ryan, Mark Doffman, Miya Masaoka, Lawrence and Judy Casserley, the Lytton family, Raymond MacDonald, Vincent Müller, Antonio Chella, Michael A. Jackson, Roger Dean, and James Rouvelle.

Thanks to the friendly people I met in the JLB, and for the generous help from Lionel Montrieux, Andy Milne, Anna Xambó, Minh Tran, Nadia Pantidi, and Tom Collins.

Finally, I thank my wife, Andrea.

Related publications

The first significant public presentation of this research was at the Music, Mind, and Invention conference, convened in 2012 to honour keynote speaker Marvin Minsky in New Jersey, USA:

> Linson, A., Dobbyn, C., and Laney, R. Improvisation without representation: artificial intelligence and music. In *Proceedings of Music, Mind, and Invention: Creativity at the Intersection of Music and Computation (MMI) 2012, The College of New Jersey, USA*, 2012a.

Methodological considerations of this research were presented at the 2012 International Conference on Computational Creativity in Dublin, Ireland:

> Linson, A., Dobbyn, C., and Laney, R. Critical issues in evaluating freely improvising interactive music systems. In Maher, M., Hammond, K., Pease, A., Pérez y Pérez, R., Ventura, D., and Wiggins, G., editors, *Proceedings of the Third International Conference on Computational Creativity (ICCC)*, pp. 145–149, 2012b.

Relevant research on interactivity was presented in the context of the 2012 Alan Turing Year in Birmingham, UK:

> Linson, A., Dobbyn, C., and Laney, R. Interactive intelligence: Behaviour-based AI, musical HCI and the Turing test. In Müller, V. and Ayesh, A., editors, *Proceedings of the AISB (Society for the Study of Artificial Intelligence and Simulation of Behaviour) and IACAP (International Association for Computing and Philosophy) World Congress, Alan Turing Year 2012, Birmingham, UK*, pp. 16–19, 2012c.

Details pertaining to the theoretical background of this research were presented at the 2013 conference of the Biologically Inspired Cognitive Architecture Society in Palermo, Italy:

> Linson, A., Dobbyn, C., and Laney, R. A parsimonious cognitive architecture for human-computer interactive musical free improvisation. In *Advances in Intelligent Systems and Computing 196: Biologically Inspired Cognitive Architectures 2012*, pp. 219–224. Berlin: Springer, 2013.

A journal article describing this research and its initial empirical results is currently under peer review:

Linson, A., Dobbyn, C., Lewis, G., and Laney, R. An artificial agent for collaborative free improvisation. *(submitted)*, in peer review.

Some of the philosophical background to this research was presented at the 2011 International Computer Music Conference:

Linson, A. Unnecessary constraints: A challenge to some assumptions of digital musical instrument design. In *Proceedings of the International Computer Music Conference (ICMC) 2011, University of Huddersfield, UK*, pp. 421–424, 2011.

And, finally, some of the philosophical implications of this research have been considered here:

Linson, A. The expressive stance: Intentionality, expression, and machine art. *International Journal of Machine Consciousness*, 5(2):195–216, 2013.

Supplementary audio

Audio material related to this thesis is available at:

`http://percent-s.com/linson-phd-thesis-audio`

Contents

List of Figures

Chapter 1

Introduction

Music may be understood in a vast number of ways, but some forms of music lend themselves especially well to traditional music theoretical analysis. Such analysis may be one of the primary means of representing specific musical traditions, for example, in terms of music theoretical rules, conventions, and even particular psychological experiences that can be tied to historically developed formal musical structures.

Recently, however, musicologists have begun to analyse different dimensions of music that are not intelligible according to the logic of traditional music theory. Social psychology, for example, has been a means of exploring social relations inherent in musical practices. Another approach has been cognitive modelling, which may involve the symbolic representations of traditional music theory, but may also use subsymbolic methods. These perspectives are increasingly being applied to the study of musical traditions that were previously only considered in terms of formal music theory.

Free improvisation poses difficulties for traditional music analysis, as it resists a definition in terms of formal rules. On the other hand, free improvisation seems to be especially receptive to the descriptive framework of social psychology, which places more significance on behaviour and interaction. Yet computational research on music often focuses on explicit and implicit musical knowledge. If free improvisation is better understood in terms of behaviour and interaction rather than in terms of musical knowledge, any approach to computationally modelling the practice must take this into account.

The notion of artificial intelligence (AI), as its name transparently indicates, is oriented towards intelligence, a concept that remains ill-defined. Historically, AI has been preoccupied with knowledge, even in its early integration with robotics. The arrival of behaviour-based robotics marked a significant shift in its orientation, away from a central focus on knowledge *per se*, and towards ideas about behaviour and interaction. Though superficially unrelated to this development, the study of human–computer interaction (HCI) has also reoriented some computing research towards behaviour and interaction, often in psychological terms.

With the development of these research perspectives, it has become possible to investigate free improvisation with a novel interdisciplinary approach. By using behaviour-based robotics and psychological research to study free improvisation, we can enquire into the cognitive underpinnings that enable free improvisors to integrate perception and action in a sociomusical context. The particular theorisation of perception, action, and context by ecological psychology, and its notion of an environment that is both natural and social, plays a crucial role in this research.

The theoretical basis for this investigation of the cognitive foundations of collaborative free improvisation is presented in this chapter. After a brief historical account of freely improvised music, the chapter introduces issues pertaining to musicology, cognition, and computation. Research questions are then presented, and an overview of the remaining chapters is given.

1.1 Musical free improvisation

Free improvisation can be regarded as an established musical practice, recognised internationally through public performances and festivals, recordings and labels, journalism and academic research. Generally speaking, free improvisation developed in the mid-twentieth-century from the activities of virtuosic instrumentalists, many of whom are still active. These musicians, typically experienced in the complex musical practices of the time, especially jazz but also classical music, often explored 'extended' techniques that pushed the limits of the mechanisms, timbres, and relation to musical notation historically associated with their instruments. Some developed and introduced acoustic, electronic, or electro-acoustic non-traditional instruments that in many cases stand in a unique relation to conventional musical practice.

As is often the case with musical and artistic practices, there is no widely agreed upon definitive, monolithic history of how free improvisation developed. The mid-1960s is viewed as a crucial period in the history of contemporary free improvisation. Musicians whose activities gave rise to the practice include those from British groups such as AMM, Spontaneous Music Ensemble, and the Music Improvisation Company; American musicians including but not limited to members of the Association for the Advancement of Creative Musicians; a wide range of European nationals, many of whom participated in the Globe Unity Orchestra; and Europe-based American expatriates such as those in Musica Elettronica Viva. A significant number of international collaborations that contributed to establishing free improvisation as a practice in its own right involved musicians active in these groups, performing at events in cities such as Baden-Baden in southern Germany; Pisa, Italy; and London, England.

Prior to the more visible mid-1960s emergence of free improvisation, a 1949 recording documents freely improvised ensemble sessions led by pianist Lennie Tristano.[1] In addition, improvised music that is widely known as “free jazz” reached international audiences in 1959 through the release of an album by Ornette Coleman,[2] and continued to rise in popularity throughout the early 1960s with the activities of Cecil Taylor, John Coltrane, Albert Ayler, and many others; around the same time, a contrasting practice of “free music” appeared in the UK with the work of saxophonist Joe Harriott and his associates (Lewis 2008, pp. 37–43). This improvised music of the early 1960s is typically perceived to have an audible relationship to the African American jazz tradition of the early-to-mid Twentieth Century, but it does not adhere to some historically conventional constraints pertaining to melody, harmony, meter, tempo, timbre, and the roles of musicians within an ensemble; it can thus also be considered a precursor of what is generally known as free improvisation.[3] It should also be noted that after the mid-1960s proliferation of free improvisation, in the early 1980s, a surge of free improvisation in Downtown New York became an important touchstone for future practitioners.

From a formal perspective, a distinguishing characteristic of free improvisation is the fact that no over-arching framework of formal musical conventions defines it. As Lewis (2004) writes, the events in Baden-Baden in the late 1960s and early 1970s, “supported by the state-owned Südwestfunk radio network, and organised by the important critic and radio producer, Joachim Ernst Berendt [...] exemplified the core conception of placing musicians in a space with few or no externally imposed preconditions — or rather, the histories and personalities of the musicians themselves constituted the primary preconditions”. Thus, the practice centres around the rapid, dynamic mutual negotiation of sound, produced without any agreed upon formal structure. To a listener, the sound can be thought of as *unstructured*, in the sense that, within a given performance, future sounds do not have a predesignated structural role, which is consistent with (though

[1]Saxophonist Lee Konitz, who participated in the sessions, recalls that “in our rehearsing together in the Tristano group, once in a while we got into a nice situation after playing those intricate lines and — I don’t remember exactly how this happened — Tristano said, ‘Let’s just improvise freely.’ [...] We went into the studio [in May 1949] intending to play things that we had rehearsed quite a lot [...]. When we finished recording those tunes, Tristano suggested we try doing one of the free improvisations that we had tried occasionally. [...] I think that’s the first recording of totally free playing. As has been stated, the only thing that was planned was the order of entry”(Hamilton 2007, pp. 210–211).

[2]Ornette Coleman, *The Shape of Jazz to Come* (Atlantic 1317).

[3]The relationship between the terms “free jazz” and “free improvisation” has a complex history that in part relates to broader relationships between national cultures and other ideological issues pertaining to race, class, and aesthetics. See Lewis 2008, pp. 247–254.

not equivalent to) the idea of a dynamically emergent structure. (This particular sense of unstructured as lacking *predesignated* structure also applies to the notions of unstructured input and unstructured environments that appear in the below discussion.)

Arising from the particular historical circumstances surrounding free improvisation, an international community developed with certain implicit sensibilities. For instance, freely improvised music typically avoids obvious extended associations with genre or style, which often means eschewing certain melodic and harmonic conventions. However, it would not be accurate to describe these sensibilities as 'rules', in the sense of traditional rules of harmony, for example. Even positively defined commonalities among practising free improvisors such as attentive listening might function as pedagogical directives or aesthetic criteria, but cannot be understood as rules in the traditional sense.

One complexity in relating free improvisation to other musical practices is that it can in principle only be identified by its performance process. That is, it is impossible to analyse a recording and determine definitively whether or not the music was freely improvised, as it is conceivable that it was performed according to a score (Lehmann and Kopiez 2010). A further complexity is that free improvisation resists a universal sonic characterisation: it may in some cases consist of readily identifiable melodic and harmonic interrelationships, or "incremental manipulations of an ongoing texture of non-referential sound", or some combination thereof; it may contain "scattered and unpredictable interruptions of silence" or may lack any breaks whatsoever; other examples may include "electronic feedback or randomly generated electronic signals with no recognisable reference to a shared tempo or harmonic structure" (MacDonald et al. 2011, p. 243).

1.2 Notation and formalisation

Rather than a sonic definition, or one rooted in conventional musical formalisms, current musicological research has begun to focus on the social dimension of music, including improvisation (see Clarke and Cook 2004). A doctoral study by Sansom (1997) has as its specific focus the psychosocial dynamics that underpin collaborative free improvisation. One of the key points of Sansom's analysis is that real-time social interactions (in addition to musical structure-based interactions) serve to guide the musical activities of co-performers in the service of elaborating musical meaning. This view of musical meaning is opposed to those rooted in traditional musical notation, where a finalised musical structure is viewed as the quintessential object of formal analysis.

Narrowly formalistic musical analysis largely developed in relation to composed music, and has carried over into the analysis of improvisation through the use of transcriptions. As a consequence, transcribed improvisations are often treated analytically as compositions. Derek Bailey (1993 [1980], p. 15), in his seminal ethnographic study of improvised music (including free improvisation), addresses this issue when he states that "transcription might help to establish matters to do with style or material used but those elements which are peculiar to improvisation and to nothing else cannot be documented in this way. [...] When the object of examination is improvisation, transcription, whatever its accuracy, serves only as a misrepresentation". The misrepresentation in part stems from the fact that the criteria for analysing composed music are closely interrelated with the practical principles of composing music, while the traditional practice of composing music differs fundamentally from the practice of improvisation, for example, in the manner described by Sansom (1997).

With respect to the relationship between music and computing, computers are a powerful means to manipulate quantitative data that may symbolically represent musical parameters of traditional notation such as pitch and duration. Given that computers lend themselves especially well to this formalistic approach to music, there is an extensive tradition of computer-aided composition (see, e.g., Hiller and Isaacson 1958; Hiller 1981; Xenakis 1992). In theory, the history of computer-aided musical composition and analysis dates back to at least 1843, when Ada Lovelace supposed that Charles Babbage's Analytical Engine, a design for a mechanical computer, could be programmed to compose music (Menabrea and Lovelace 1843, Note A).

Over a century later, Alan Turing (1950) and Claude Shannon (1950), key figures in the history and development of AI, made marginal references to the potential for computers to play a role in musical composition or orchestration. Another key figure in AI, Herbert Simon, stated in 1957 that practical work in musical AI was already underway (McCorduck 1979). He was referring in particular to work by Lejaren Hiller and Leonard Isaacson on the *Illiac Suite for String Quartet*, which began as early as 1956 and used integers to represent musical notes (see Newell et al. 1958; Hiller 1981).

The dominant tradition of AI research to which Simon belonged aimed to develop human mental competence by a machine. This research paradigm was historically linked to empirical psychology, with both psychological and AI research seeking to provide evidence about how the human mind works (Newell and Simon 1976). Within the field of AI, the prevailing trend of empirical research sought to establish that the human mind operates according to a unified calculus of symbolic representation and manipulation, an "information processing" model of the mind. The basis for this view, now considered classical, can be summarised with two hypotheses presented by Newell and Simon (1976):

(1) "the necessary and sufficient means for general intelligent action" can be provided by a complete and closed system of symbols, linked into structures that correspond to objects, and of processes of "creation, modification, reproduction and destruction" that obey physical laws (p. 116),

and

(2) that such a system "exercises its intelligence in problem solving by search — that is, by generating and progressively modifying symbol structures until it produces a solution structure" (p. 120).

In the specific case of the *Illiac Suite for String Quartet*, Hiller and Isaacson (1958) used a digital computer programmed with "mathematical operations which express the rules of composition" for studies of "strict counterpoint", "dissonant chromatic writing and tone-row generation", and "writing music by more abstract procedures based upon certain techniques of probability", namely, Markov chains. These techniques relate to early AI research on "human mental functioning", especially problem solving (explicitly linked to a discussion of music and creativity in Newell et al. 1958; see also Simon and Newell 1962). Computer systems for musical improvisation, discussed in the next chapter, would not appear until the late 1980s.

1.3 Situated activity

Computer-aided quantitative methods for compositional analysis and generation, including but not limited to the use of traditional notation, have, due to their success, influenced research on improvisation. But, as Clarke (2004) points out, a quantitative approach "entirely misses the social dimension of performance", including the "interactions between performers" (p. 91). Elsewhere, he writes that ensemble improvisation "must be understood as an interactive process between performers". In particular, for "free group improvisation [...] the guiding principles or constraints may be primarily concerned with the kinds of interactions between players [...] rather than the selection and filtering of the material itself. This emphasises the importance of viewing improvisation as a social process" (Clarke 1992, p. 790). The "selection and filtering" of material to which Clarke refers is one of the predominant approaches to investigating improvisation using computational models, often presented as explanatory models of human cognition (e.g., Johnson-Laird 2002).

Concerning the investigation of human cognition, Clarke (2005b) points out that "explicitly psychological research on improvisation [...] has focused on the productional features of improvisation, largely ignoring questions of perception and the relationships between co-performers" (p. 175). Productional features in this sense refer to accounts of why one note is played rather than another, for instance, due to harmonic rules, probabilistic predictions of note successions, and other forms of rational calculation. He goes on to say that "similarly, the treatment of the musical material has itself [...] been narrow, and has often treated improvisation as if it was a special case of musical problem solving" (p. 175). This criticism of the research focus on musical *production* in improvisation echoes the criticism, noted above, of the use of compositional criteria in the *evaluation* of improvisation. Clarke concludes by stating that "it is in the social character of improvisation that psychological research still has much to explore" (p. 175). One aim of the present study is to address this gap in research, and to present an alternative to existing production models and evaluation methods for research on improvised music.

Clarke's (2005b) criticism of the notion of "musical problem solving" in improvisation is suggestive of the general dichotomy between the formal (i.e., compositional, notational) and socially oriented approaches to understanding music. A similar dichotomy is also evident in the history of AI. As described above, the predominant view of (natural and artificial) intelligence, found in theoretical and technological models of cognition from the 1950s to the 1970s, was a formal-symbolic, information processing view. Cognitive processes were thought to be rooted in internal symbolic representations of an external world, subject to formal operations and constraints.

In the 1970s, this formal-symbolic view was challenged by research that regarded environments or situations as co-determiners of the thoughts and actions of individual agents. A broad theoretical critique of AI by Hubert Dreyfus (1992 [1972]) was instrumental to this challenge. His views were directly taken up by a number of empirical AI researchers whose subsequent work became influential in the field (e.g., Winograd and Flores 1986). According to Dreyfus (2008), this shift in practice facilitated — at least indirectly, if not directly — the advent of behaviour-based AI, including the robotics research of Rodney Brooks, a main topic of the present research, discussed in further detail below.

Arguably, Brooks' approach to dynamic agent–environment interaction implicitly emerged not only from the AI research climate that followed Dreyfus' critique, but also from research in ethology that developed from the insights of James Gibson (Kirsh 1991). Already in the 1960s, both the biology-oriented psychology research of Gibson and the phenomenological philosophy of Martin Heidegger and Maurice Merleau-Ponty began to call attention to the relation between an environment or situation and humans' ability to make sense of and act in their world. Despite key differences among these thinkers, their ideas have much in common, especially their emphasis on the notion that subjective determinations of significance are made in a shared context of historical, sociocultural contingencies. Dreyfus' (1992 [1972]) critique of the information-processing view of intelligence that was taken up in AI research largely drew upon phenomenology, especially that of Heidegger and Merleau-Ponty. However, in a parallel development, Gibsonian or ecological psychology became increasingly influential in ethology, also as a critical response to a predominant information-processing view of biological perceptual systems.

The idea that an environment or situation co-determines individual thought and action is expressed by the notion of 'situated activity'. A popular example in the AI literature, in the sub-field known as situated robotics, is Maja Mataric's (1990) boundary-following robot (an implementation of Brooks' design principles). The robot exhibits, for example, wall-following behaviour, despite no explicit programming about what constitutes a wall. Instead, the dynamic interplay of a robotic sensor-based perceptual system and a collection of actuators that drive certain patterns of activity gives rise to meaningful behaviour.

Another form of situated activity, in this case, related to musical improvisation, is described in a widely cited auto-ethnographic account by Sudnow (2001). Referring to Sudnow's account, Horst Hendriks-Jansen (1996) observes that "there is a crucial difference between the 'middle F' used in the explicit notation that serves as part of the scaffolding in pedagogy and the 'meaning' of this key as it emerges from interactive experience" (p. 314). This description reinforces the above accounts of improvisation with respect to the problematic role of conventional notation in understanding musical meaning.

Linking the above two examples from music and robotics, Hendriks-Jansen underscores the shortcomings of the formal-symbolic view of thought and action: "Both the F struck by the pianist and a specific instance of wall-following by Mataric's robot are objectively recognizable and may be categorised by such labels as 'middle F' and 'left wall,' but an objective classification along these lines does not provide the basis for an explanation of the sensorimotor mechanisms involved" (p. 314). In contrast, a situated perspective seeks to address why such labels do not reflect the constantly shifting emergent relationships between an agent and an environment.

1.4 Interaction

An imagined objective viewpoint implies immutable object designations, such that, for instance, a chair is in all circumstances well described as a 'chair'. From a situated perspective, however, relationships between an agent and an environment are structured by the link between perception, action, and meaning. Gibson (1966) conceptualises this link with the idea of *affordances*, which refer to potential significance in an environment, relative to an agent's capacities or needs. Thus, affordances differ from abstract physical properties or universal designations, such that a chair may only be relevant in some circumstances if it is 'sit-on-able', which may only be the case if it is the right height (Gibson 1979).

The concept of affordance is further articulated with another example: to a water bug, a body of water appears as a surface that "affords support", but not to a heavy terrestrial animal (Gibson 1979), although, for the latter, it may afford swimming or boating. Dreyfus (2008) explicitly connects the theories of Gibson and Merleau-Ponty; for both, "how we directly pick up significance and improve our sensitivity to relevance depends on our responding to what is significant for us" in a given context (p. 361). In other words, our detection of and responses to affordances may change from moment to moment, relative to both natural and cultural influences (Gibson 1966).

A sustained exploration of the link between perception, action, and meaning with respect to music is given by Eric Clarke's (2005a) "ecological approach to the perception of musical meaning", an explicitly Gibsonian investigation of music cognition that includes the topic of improvisation. In the context of affordances, Clarke points out how in the everyday activity of listening, sound may be differently perceived and thus afford different actions, depending on both natural and social contexts. Although he covers a wide range of examples, he notes that musical "improvising involves a kind of listening-while-performing that highlights the relationship between perception and action [...] in a particularly acute manner" (p. 152). In improvisation, listening and playing are inextricably intertwined with each other and with a socially elaborated musical context.

By viewing musical improvisation in ecological terms, it can be said that through sound, co-performers afford different opportunities to one another, in that they can respond to or interact with sound in a variety of ways. This results in a dynamic interrelationship that can be understood as a collaborative performance. Andy Clark (1997) describes this kind of interrelationship as one of 'continuous reciprocal causation'. Notably, although Clark introduces this concept in a broader discussion of agent–environment relations (also referring to Brooks, Gibson, and Merleau-Ponty, among others), to explain it, he draws the following analogy: "The players in a jazz trio, when improvising, are immersed in [...] a web of causal complexity. Each member's playing is continually responsive to the others' and at the same time exerts its own modulatory force" (p. 165). His description is concerned with the emergent properties of distinct, yet coupled components. Following his analogy, a collaboratively improvised performance emerges from the mutually influential activities of the participant co-performers.

Conversely, without such 'continuous reciprocal causation', a meaningfully collaborative improvisation does not emerge. Using the simpler example of a duet, if one improviser's actions have no effect on the other's, neither participant is likely to perceive the performance as collaborative (barring some coordination external to the performance). Moreover, in this situation, the affordances offered by a non-responsive player would progressively lose significance: if one performer does not perceive any effects of their playing on the other, the impetus to further action is diminished.

A similar connection between interaction, affordance, and responsiveness is described by Donald Norman (1999), who first introduced the concept of affordances into the discourse on HCI design: "Although all screens within reaching distance afford touching, only some can detect the touch and respond to it. Thus, if the display does not have a touch-sensitive screen, the screen still affords touching, but it has no effect on the computer system" (p. 39).[4] In this scenario, if someone sought to complete a computer-based task by touching the screen, and the screen proved to be non-responsive, he or she would be expected to seek out another input device; the affordance is still there, but it is no longer perceived as significant in the course of certain activities. In other words, affordances may lose their significance relative to a given situation, as with the sounds of a non-responsive improvisor.

Given the above-mentioned confluences between ecological theory and phenomenology, it is perhaps not surprising that the latter has also influenced HCI. Drawing on the phenomenological tradition, Paul Dourish's (2001) work on embodied HCI states that "the key feature of interaction with computation is how we *act through* it to achieve effects in the world" (p. 137, original emphasis). His view relates closely to what Andy Clark and David Chalmers (1998) call the "extended mind". This refers to the idea of a tight coupling between an agent and an artefact that makes certain actions — including, or perhaps especially, cognition — resist being defined as taking place within the apparent physical boundaries of the agent. For example, some actions, such as performing long division with a pencil and paper, must be viewed as part of a continuous reciprocal causation with objects and processes in the environment, not as confined to an individual brain or body.

Along similar lines, Sudnow (2001) relates that once he developed significant skill as an improvising pianist, he was able to sense that his hands or fingers were the ones doing

[4]Norman stresses the difference between 'real' and 'perceived' affordances and their relation to 'convention', but Clarke's (2005a) formulation of essentially the same three concepts is more elegant. Instead of drawing these category distinctions, Clarke refers to how 'remote' a particular action-oriented significance is, regardless of whether this remoteness is influenced by physical reality or cultural convention. To illustrate his point, Clarke uses the example of a violin, which although it technically affords burning, "social factors ensure that this is a rather remote affordance"; under ordinary circumstances, "the musical context regulates its affordances" (p. 38).

the thinking or music-making, such that the piano became an embodied extension of his agency. Two ideas that relate to the above discussion of perception, action, and meaning can be distinguished here. The first is the tightly coupled perception-action loop that links the piano keys with the body as the result of progressive familiarisation through physical practice. The second is the notion of meaningful action: acting *through* the piano to achieve effects in the world; acting in such a way as to give 'middle F' a meaning that depends on the context in which it was played — a context that is as much socially determined as it is physically. For Gibson (1966), the world is subjectively intelligible as a web of meaning constituted through both natural and sociocultural forces.

1.5 The Physical and the Social

The physical tight coupling of a performer and a musical instrument can be viewed as a single system and, despite a rich social context, their physical interaction dynamics could in principle be effectively described by a mathematical model. While this would not address the outcomes of specific performances, it could reveal an arguably relevant dimension of playing an instrument, for example, at the nexus of physiological and material constraints. However, when multiple performer-instrument 'systems' are engaged in collaborative improvisation — i.e., "each member's playing is continually responsive to the others' and at the same time exerts its own modulatory force" (Clark 1997, p. 165) — a purely physical level of description fails to address the relevant features of the situation. As described above, the musical interactions in collaborative improvisation are socially situated, such that a performance emerges from a social dynamic embodied in the medium of sound.

In the context of improvising, there is a different significance to sound from one's own instrument and sound from another co-performer's instrument. Taking current psychological research into account (discussed in the next chapter), it appears that intentional agents have the capacity to recognise one another by perceiving behaviours against the background of an environment. This recognition is facilitated by observation and interaction, which reveal an interplay of relevant perceptual cues that indicate the presence of others. When other intentional agents are present, it makes possible the experience of participating in a shared situation, especially through mutual engagement in a coordinated activity. Thus, while the processes of sonic production and reception (e.g., of a constant tone on a musical instrument) can be described by relatively context-independent physical and psychoacoustic models, these processes are substantially modified in a social context comprised of other individuals, such as a musical interaction or verbal conversation.

For investigations of social or sociocultural phenomena, the relevant significance is not inherently physical, though it may involve physical aspects (e.g., Givan 2009). As Bailey (1993 [1980]) states, with respect to the study of improvisation, "even when man's senses are supplemented by such devices as the oscillator and the frequency analyser the result is only a more exact picture of the irrelevancies. It still has nothing to say about the forces behind the music making" (p. 15). More generally, as Martyn Hammersley (2010) stresses, any research methodology should "ensure that the assumptions built into measurement procedures correspond to the structure of the phenomena being measured" (p. 420). In this respect, for any research, the methodology and theoretical standpoint must be understood as closely interrelated. For the present research, this interrelationship is concretised by the way improvisation has been theorised above, which lends itself to a qualitative investigation of collaborative experience.

1.6 Subsumption architecture

The relation between background theory and research methodology also arises in the aforementioned dichotomy between formal-symbolic AI and Brooks' behaviour-based approach. In discussing this contrast, Brooks (1991) described a new way of formulating intelligence: It should "be reactive to dynamic aspects of the environment", and "be able to generate robust behaviour in the face of uncertain sensors, an unpredictable environment, and a changing world" (p. 1229). He proposed to address these requirements by a set of robot design principles he developed called the 'subsumption architecture' (Subsumption, hereafter), "which deliberately changed the modularity from the traditional AI approach". In particular, this meant a new way of conceiving of how to combine computational subsystems or 'modules', and how to connect them to sensor input and actuator output.

One key feature of Subsumption is "short connections between sensors and actuators, making it plausible, in principle at least, to respond quickly to changes in the world" (Brooks 1991, p. 1230). Another key feature is that such connections are organised into 'layers', that is, networks of modules that correspond to behavioural traits. In contrast to a horizontal decomposition, in which a sequence of modules forms a long chain between sensor input and actuator output — generically referred to as a 'sense–model–plan–act' framework — Brooks' describes his innovative arrangement of layers as a vertical decomposition.

Vertically decomposed behavioural layers are organised to interact with and influence one another by the asynchronous transmission of simple messages (i.e., with a low number of bits). These messages intervene in the dataflow from sensor to actuator in other layers, by 'suppressing' (masquerading as) local sensor values or 'inhibiting' (preventing) data from reaching an actuator. This approach to modularity is robust, because a failure in one module does not necessarily result in complete system failure, in contrast to a traditional horizontal decomposition. Also, the debugging process is comparatively straightforward, where shortcomings discovered during deployment immediately suggest an association with a specific layer, and thus need not be traced through the entire system.

In more general terms, this design strategy lends itself to rapid prototyping, with progressive fine tuning facilitated by real world deployment. Thus, in addition to the other differences between the Subsumption approach and that of traditional AI described above, the development process also differs greatly, considering that the latter typically relied on idealised laboratory conditions (e.g., SRI's *Shakey*), or the further idealised conditions of software-based simulations of agents and environments (e.g., Winograd's *SHRDLU*).[5] In light of this contrast between real and ideal conditions, Brooks points out that the evaluation criteria associated with traditional AI are not applicable to robots built using Subsumption.

As Brooks (1991, p. 1232) states, Subsumption robots "operate in much more uncertain, and much more coarsely described worlds than traditional AI systems operating in simulated, imagined worlds. The new [Subsumption] systems can therefore seem to have much more limited abilities". Moreover, these abilities cannot be appropriately evaluated according to the static results of a terminating causal chain, as with the 'sense–model–plan–act' conception of intelligent behaviour. With the Subsumption conception of the agent–environment relationship, an appropriate evaluation must take into account the notion of interactively emergent behaviour, which entails a continuous dynamic view of system performance.

1.7 Research questions

This chapter has presented collaborative musical free improvisation as a situated activity, more aptly characterised as a socially-constituted qualitative experience than as a quantitatively measurable phenomenon. A collaborative performance has been described as one that dynamically emerges from the rapid negotiation of unstructured sound from mutually responsive participants. In considering this emergent phenomenon, the notions of tight coupling and continuous reciprocal causation may describe both agent–environment and agent–agent relations. These themes relate closely to aspects of

[5]Although Winograd's later work was influenced by Dreyfus, *SHRDLU* pre-dates Dreyfus' (1992 [1972]) critique of AI.

Subsumption also described in this chapter. It is worth noting that Brooks' graduate student, Jonathan Connell — whose work led to important revisions in the Subsumption specification — describes a (non-musical) mobile robot he developed as having the ability to "improvise", referring to its traversal of dynamic physical environments such as busy offices (Connell 1989).[6] In light of these connections, a general research question of this thesis is, *Can an artificial agent built using Subsumption effectively perform human–computer interactive collaborative free improvisation?*

This agent, a Subsumption system called *Odessa*, is used to investigate a further set of three research questions, based on research in psychology and HCI described above. These research questions (numbered points) are presented here with some relevant context (bullet points):

- Within the sonic environment of collaborative free improvisation, an interplay of certain observable behavioural cues can signify the presence of other intentional agents. Thus, for a human performing a collaborative free improvisation with *Odessa*,

  (1) Will the system be perceived as an intentional agent by (sonically or musically) exhibiting the appropriate behavioural cues?

- Assuming an intentional agent is perceived, it can participate in a collaboration with others if it behaves competently with respect to the collaborative practice (in this case, free improvisation), and it is responsive to the other participants in the collaboration. Thus, for a human performing a collaborative free improvisation with *Odessa*,

  (2) Will the performance be experienced as collaborative if the system demonstrates competence as a free improvisor and is responsive to the human co-performer?

- Assuming an intentional agent is a competent and responsive free improvisor, if it contributes to the collaborative sonic environment with affordances that present multiple possibilities for action, other co-performers can engage in the collaboration by pursuing these possibilities. Thus, for a human performing a collaborative free improvisation with *Odessa*,

  (3) Will the collaboration be engaging to the human co-performer if the system provides sufficient musical affordances for the human co-performer?

[6]Another student of Brooks, Ian Horswill, contributed to a principled consideration of everyday improvisation (such as cooking breakfast) in terms of agent–environment dynamics, through the lens of AI research (Agre and Horswill 1992).

1.7.1 Preliminary objections

Regarding the use of Subsumption for this research, it can be noted that Brooks' robots are based in embodied physical mechanics, which, in principle, can be adequately described with mathematical models. This seems at odds with the socially oriented account of improvisation presented above. Indeed, as Lucy Suchman (2007, p. 15) notes, for Brooks, situatedness is "evacuated of sociality, at least as other than a further elaboration of an environment understood primarily in physical terms". Yet Brooks' criticism of formal-symbolic information-processing AI does appear to resonate with the above-presented criticisms of traditional musicological approaches to improvisation and their supporting cognitive models. Following this line of inquiry, in relation to *Odessa*, Suchman's concern does not apply. On the contrary, from a methodological perspective, *I propose that a Subsumption agent for human–computer collaborative free improvisation can be qualitatively investigated on the basis of the human interaction experience.* In the present context, this investigation has the potential to offer insights into the social processes that underpin collaborative improvisation between humans.

One notable consequence of Subsumption is that it obviates the need for centrally planned action and, perhaps more significantly, precludes a central repository of stored representational knowledge. Thus, any computer agent built with this approach will be, by design, limited in its capacity to respond to cultural references, and it will lack a shared sense of long-term emergent musical structures and conventional musical knowledge.[7] Although these limitations may detract from other forms of musical performance, it is assumed that they do not pose a grave problem for free improvisation. In its 'real world' contemporary performance context, it is possible for co-performers to exhibit mutual responsiveness on the basis of short-term sonic gestures, and performances are amenable to discontinuous, sudden textural shifts. This supports the initial general proposal that Subsumption is well suited to the design of the proposed agent.

1.8 Overview and contributions

In sum, research on free improvisation has been conducted from the perspectives of ethnomusicology (e.g., Bailey 1993 [1980]) and social psychology (e.g., Sansom 1997), but without a specific focus on its constitutive cognitive processes. The present study

[7] This is not to deny, however, that "technological inventions [...] are fundamentally human (and social) constructions, and as such embody and enable specific values, agendas, and possibilities" (Ensmenger 2011, p. 26); see also Lewis (2000, p. 33): "As notions about the nature and function of music become embedded into the structure of software-based musical systems and compositions, interactions with these systems tend to reveal characteristics of the community of thought and culture that produced them".

is concerned with real-time cognition in improvisation. The orientation of this study, that of ecological psychology, is drawn from Clarke (2005a), where improvisation serves as a key example in his elaboration of an ecological theory of music perception and meaning-making (see also Clarke 1992, 2005b).

Broadly speaking, Gibsonian psychology theorises a relation between agent cognition and the environment, with the environment defined in both natural and social terms. This view of agent–environment dynamics, however, does not specifically account for identifying other agents within an environment by their behaviour. A compatible account of how we recognise other agents is given in empirical psychological research on the attribution of intentionality, described in terms of observing an interplay of specific behavioural cues. A key starting point for the present research is formed by connecting concepts from cognitive ethology (the perception–action cycle, agent–environment dynamics), and human developmental psychology (how we make intentional attributions by interpreting a specific interplay of perceptual cues), and relating these to Clarke's ecological conception of music.

Another key starting point is a connection between agent–environment dynamics, the interplay of behaviours, and Subsumption robotics, and the relation of these to improvisation. Subsumption robots use an interplay of low-level mechanisms and behaviours to give rise to higher-level behaviours. Such robot agents are particularly flexible and robust in the face of dynamic, ever-changing environmental conditions, which suggests the concept of improvisation. An important aspect of freely improvised music is that performers remain flexible in the face of dynamic, ever-changing conditions in the musical environment.

Following these theoretical connections, I designed a Subsumption agent, a parsimonious dynamic model from which complex improvisational behaviour can emerge. To gauge the system's ability to competently perform free improvisation, it was deployed as a co-performer with expert human improvisors. These human experts judged the global behaviour of the system after participating in collaborations with it. Their largely positive judgements indicated that competent improvisational behaviour is facilitated by the underlying collection of simple mechanisms. The system's behavioural mechanisms are not rooted in neuroanatomy, so the study is not intended to validate a biomimetic model. Rather, the study demonstrates that complex improvisational musical behaviour may be underpinned by a small collection of simple mechanisms, as opposed to a large collection, or one consisting of more complex mechanisms.

The empirical investigation proceeded in two stages. The first stage consisted of eight cases studies using an initial prototype for duet collaborations with human experts. Their experiences were documented and analysed, and a second iteration of the system was developed on the basis of these results. The second study was smaller in scale, but demonstrates the impact of the updated model.

The findings indicated that a Subsumption design was fit for the purpose of the agent, and that the qualitative approach to analysing the participant experience was fit for the purpose of the study. With respect to the three more detailed research questions above, on the basis of participant feedback, the findings indicated that the interplay of appropriate behavioural cues typically led to the attribution of intentionality; the performances with system were frequently perceived as collaborative on the basis of the system competence and responsiveness; and the collaborations with the system were regarded as engaging when the affordances were sufficient, although the affordances were not always sufficient. This latter issue will be addressed in future work.

The research presented here is the first experimental study of cognition in free improvisation based on an ecological model of music perception. In short, I developed a behaviour-based artificial agent that improvises with human musicians; the system was deliberately limited in order to focus the investigation on particular aspects of cognition. With this approach, I demonstrated that cognitive architecture plays a greater role in free improvisation than merely facilitating the adherence to explicit or implicit cultural rules, a common assumption in previous research. This investigation contributes to the empirical grounding of Clarke's ecological model, and has implications for music cognition and improvisation studies.

This research also contributes to the field of Subsumption robotics, as a significantly new application of its principles to an artificial musical agent. At the same time, it is a contribution to research on artificial improvisation systems, an area in which Subsumption principles have not been previously applied in this manner. Notably, it is the first time that a Subsumption system has been used for real-time interactive music.

The remaining chapters are organised as follows: Chapter 2 surveys other systems related to this research, and reviews literature related to the remaining chapters, including relevant background on psychology and methodology. Chapter 3 documents the system design and implementation. Chapter 4 provides the methodology used for the experiments and evaluation. Chapter 5 presents the analysis of the case studies with the first prototype. Chapter 6 describes the second prototype and corresponding follow-up study and analysis. Chapter 7 summarises the research and proposes future work. Background on the implementation can be found in Appendix A, and details of the supplementary audio can be found in Appendix B.

Chapter 2

Literature Review

In the AI community, one of the most substantive explicit thematisations of interaction and its connection to agent–environment relations and the social realm was presented in 1995, in a special double volume of *Artificial Intelligence* (72/73) on the topic of *Computational Research on Interaction and Agency*.[1] The introduction by Agre (1995) notes the "many subtle and improvised ways in which people structure their actions in accord with the demands of moment-to-moment meaningful interaction", pointing out that "these phenomena may lead future computational research to rethink its basic concepts in ways that can do justice to the improvisatory nature of human action" (p. 27). Agre also envisions that future research in AI "might explore the ways in which agents can improvise their interactions with one another" (p. 41). Curiously, the entire edited collection does not refer to improvised music, despite the fact that the first artificially intelligent interactive musical improvisation systems had already been introduced at the time.

Since the introduction of such systems approximately twenty-five years ago, a variety of human–computer interactive computer systems have been designed and developed for musical improvisation (see Dean 2003). Most of those discussed in this chapter are specifically designed for free improvisation, with a few relevant exceptions. All of the systems described here can be thought of in Robert Rowe's (1993, p. 8) terminology as "player-paradigm systems", which "regard the program as implementing a musical voice which may be related to, but is still in an audible way distinct from, the performance of a human partner". This type of system is opposed to "instrument-paradigm systems", "which treat the system as an extended musical instrument".

In order to better understand *Odessa*'s architectural design and why it is significant, this chapter discusses several player-paradigm systems, including two of the original

[1] Republished as Agre and Rosenschein 1996.

interactive improvisation systems that have continued to be widely influential: Rowe's *Cypher* and George Lewis' *Voyager*, both of which use a multi-agent design and perform freely improvised music. Another free improvisation system, in fact, a group of systems, discussed here is Tim Blackwell's collection of *Swarm* systems, including *Swarm Music*, and the related *Swarm Granulator* (in collaboration with Michael Young), which use the techniques of swarm intelligence. Two other multi-agent systems are also discussed: Nicholas Collins' *Free Improvisation Simulation*, which shares some similarities to the design of *Odessa*; and the *VMMAS* by Wulfhorst, Nakayama, and Vicari which, although it does not perform free improvisation, highlights certain issues in system design and human–computer interaction that are relevant to the research on *Odessa*. In a subsequent section, a cybernetic model of human–computer interactive free improvisation is discussed, as is Braasch, et al.'s *CAIRA*, which shares some of the same theoretical underpinnings as the *Odessa* design. The chapter also discusses Joanna Bryson's *Reactive Accompanist*, which is not designed for free improvisation, but is the first Subsumption system for music.

After reviewing the systems discussed above, literature on psychology will also be discussed, to more sharply define the investigation of *Odessa*'s apparent intentionality, and its capacity to collaborate and be engaging. This will include a further discussion of ecological psychology, first introduced in the previous chapter, with respect to collaborative improvised music and the dynamics of interaction between agents. It will also include an expanded discussion, drawn primarily from developmental psychology literature, on the attribution of agency and intentionality on the basis of perceptual cues, also introduced in the previous chapter. The conclusion of this chapter will summarise the findings and indicate how they have affected the design and experimental evaluation strategy for *Odessa*.

2.1 Contrast to composition systems

In a broad sense, a central aim of both improvisation and composition is produce music, and each represents a different approach to this aim. With respect to computer music systems, generally speaking, a basic difference between composition and improvisation systems is that the former are used to produce music that can be performed at any time in the future, whereas the latter must produce music interactively in real time. Examples of composition systems include Hiller and Isaacson's (1958) custom software for the ILLIAC computer, Xenakis' (1992) UPIC, and Cope's (2005) EMI. Each are means of composing musical scores that can be performed independently of their production.

With respect to the specific use of AI in composition systems, a number of techniques and combinations thereof can be used to produce scores, such as mathematical models (e.g., of stochastic processes), symbolic knowledge-based systems of rules and constraints, tonal grammars, evolutionary methods, and machine learning using (e.g.) neural networks (Papadopoulos and Wiggins 1999). Given the processing speed of current computing systems, many of the same techniques can be used in improvisation software, and often are. Notably, Rowe's (1992) Cypher uses some of these techniques to produce and perform composed music, while also using them for real-time improvisation. Other AI-based improvisation systems also typically use one of more of these techniques, however, *Odessa* does not use any of those named above. Nevertheless, it has commonalities with other improvisation systems. These systems, and their relation to *Odessa*, are considered below.

2.2 Survey of systems relevant to the proposed research

2.2.1 *Cypher*

Robert Rowe's *Cypher* (Rowe 1992, 1993), a real-time interactive computer music system that performs improvised music with human players, is based on Marvin Minsky's (1986) influential notion of the mind as a collection of competing and cooperating agents. As Rowe (1992) describes, Cypher consists of "two major components: a listener and a player. [...] The listener classifies features in the input and their behaviour over time, sending messages that communicate this analysis to the player" (p. 43). He continues:

> Cypher coordinates the output of many small, independently functioning agents. Analytically, these agents combine to make larger agencies that perform complex tasks, such as chord identification, key identification, beat tracking, and phrase grouping. Compositionally, collections of agents can be engaged in the generation of new musical output, which is in turn modified through the preferences of a compositional critic, a separate agency devoted to the evaluation of musical structures and their evolution in time. (p. 62)

When the system is engaged in human-interactive improvisation, "the output of the computer part is also a transformation of the human performers' input; the transformations used are a function of the features extracted from the music played to the computer" (p. 61).

Although Rowe's design models global intelligence as a collection of more simple behaviours, which seems related to a Subsumption design, in contrast, *Cypher* constructs explicit internal musical representations. Moreover, its potential actions are evaluated by a deliberation mechanism, a typical feature of traditional cognitive models. The design of *Odessa* differs in these respects, as shown in the next chapter.

2.2.2 *Voyager*

George Lewis' *Voyager* (Lewis 1999, 2000) — which pre-dates *Cypher* by a few years — also reflects the notion of a diverse collection of cooperative and competing agents. In *Voyager*'s case, Lewis (2000, p. 33) explicitly characterises the system as a "virtual improvising orchestra" of 16 "players", performing in various combinations, rather than a single, overarching entity, although Lewis (1999, p. 106) does refer to the system as having a "complex global 'personality' ". He notes that "control of musical process is shared among players", and that "inter-player communication takes place without necessarily involving a central authority" (Lewis 2000, p. 37), a general characteristic of multi-agent systems.

Despite the technical and aesthetic differences between the two, another general commonality between *Voyager* and *Cypher* is that, as Lewis (1999, p. 103) describes, "musical decisions made by the computer are based on a feature extraction and development procedure, where analytic processes deposit their outputs into a block of variables which represent the state of the input at a given moment". But one significant distinguishing characteristic of *Voyager* is what Lewis refers to as "its nonmotivic approach to form" (p. 105), as opposed to the more traditional model of compositional development of melodic and harmonic material in *Cypher*. In nonmotivic improvisation, "the global aggregation of sonic information, considered in a temporal sense, is privileged over moments of linear development". In other words, in common with established practices of free jazz and free improvisation, "moment-to-moment choices can shift unpredictably", with rapid fluctuations in polyphonic density and "sudden changes of mood, tempo and orchestration" (Lewis 2000, p. 36).

Lewis (1999, p. 106) speaks of *Voyager*'s "accumulation of many small details of input and output representation" in the course of a human-computer musical interaction as being capable of signifying emotional states and intentional acts. Another way of referring to this phenomenon is with the notions of interactive emergence and situated activity (Hendriks-Jansen 1996). In other words, a system is programmed to perform certain low-level tasks, such as altering an integer stored in a buffer, in response to a change in another buffer. But, in the midst of a collaborative musical improvisation, the system's behaviour can be observed and described at a level commensurate with musical meaning, for example, as a performer initiating a change in musical mood. *Voyager*'s successful functioning in this context makes use of complex musical representation, which the design of *Odessa* deliberately lacks; the next chapter shows how the latter achieves its successful functioning using Subsumption design principles.

2.2.3 *Swarm Music and Swarm Granulator*

Tim Blackwell has designed human–computer interactive free improvisation systems using the techniques of swarm intelligence, in this case based on the premise that there is a confluence between the emergent dynamics of animal flocks or swarms, and those of a freely improvising music ensemble (Blackwell 2001; Blackwell and Bentley 2002a,b; Blackwell and Young 2004a,b; Blackwell and Jefferies 2005; Blackwell 2007). As Blackwell and Young (2004a) describe, flocks or swarms and free improvisation both exhibit higher-level patterns that are referred to as "self-organisation", where an accumulation of low-level interactions yields observable larger structures over time, even when the latter are neither deliberately built nor centrally coordinated. This approach follows from computational research on particle swarm optimisation, first introduced by Kennedy and Eberhart (1995), which was inspired by a consideration of the flocking and swarming behaviour of animals but had a stated motive of modelling human social behaviour. Blackwell and his collaborators, in their computational modelling of swarm dynamics, have devised a means of using musical input to influence the dynamic behaviour of virtual swarms, as well as mapping the swarms to musical output.

Blackwell and Young (2004a) describe one application using "multi-swarms", where "each swarm can be thought of as a separate musical individual", or, "alternatively, the multi-swarm can be compared to a single musical personality" (pp. 131–32). Thus, although swarm intelligence is a distinct technique, and the term 'agents' is used only in the more limited sense of swarm 'particles', at a higher level, *Swarm Music* and *Swarm Granulator* have some similarities to the other multi-agent interactive music systems described here. (Another system, *Swarm Tech-tiles* (Blackwell and Jefferies 2005), operates in largely the same way, but uses visual rather than sonic input to produce sonic output, and is therefore not further discussed here.)

Swarm Music maps input events from a human performer, or from other virtual swarms, to a virtual coordinate space. A swarm of virtual "particles" is then drawn to these input events, which serve as dynamic "attractors" for the swarm behaviour. When the particles are mapped to musical output, the resulting melodic streams can be heard as responses to the human performer, while also influencing the performer's musical choices that follow. In this sense, the system is portrayed as an "artificial improvisor".

Swarm Granulator uses a similar procedure, but rather than musical events, it operates on the level of timbre, extracting sinusoidal information from input and applying it to granular synthesis in output. Multi-swarm dynamics govern the input–output transformations such that "the result is a stream of audio varying from sparse and irregular bursts to highly dense clouds of sonic material, slowly or rapidly evolving in pitch and amplitude range and in timbre" (Blackwell and Young 2004a, p. 135). The authors' long-term research aim is to integrate these two swarm-based systems to "listen" to input and modulate output at the different structural levels that the systems address, simultaneously. The swarm intelligence focus on emergent phenomena is relevant to *Odessa*, although its global behaviour emerges from different kinds of low-level interactions, described in the next chapter.

2.2.4 *Free Improvisation Simulation*

Another multi-agent system, Nicholas Collins' *Free Improvisation Simulation* (2006), is an exploration of a group of agents that model "human tendencies" in their behaviours, including their playing style, their responses to each other, and their responses to a human performer. As Collins describes, the "independent agents each with varying behaviour give some interesting complex responses from relatively simple code" (p. 182). The manner in which the agents are "cross-connected" is such that pitch and onset information from each agent, and from the human performer, can affect the behaviour of each agent. In particular, each agent has a fluctuating internal state that determines whether it uses its unique, fixed generative algorithm to produce output, or whether it uses the output from other agents, or from the human performer, as a source for its own output.

Collins (2006) also ascribes human behavioural attributes, such as "taciturnity", "sloppiness", "keenness", and "shyness", to a number of parameters. For example, "taciturnity" describes "onset detection threshold". He also refers to the notion of emergence, to describe the sustained, coherent performance of freely improvised music despite the relative lack of "a priori style grounding", that is, extensive rules for representing a given style. In his words, "there is no long-term control over the values taken on; a succession of abrupt parameter space jumps is assumed to cause enough interest, in the spirit of promoting musically varied behaviour" (p. 182). Although his system uses "melodic cells", that is, precoded musical motifs for each agent, and a gradually changing memory of the incoming human musical motifs, as he puts it, "this is somewhat blurred by the short motif sizes and flurries of activity" brought about the by rapid state changes.

The behaviour-determining parameters of the agents are configured in advance of each performance, which significantly affects their musical personalities. With *Odessa*, however, not only is a different system architecture used, but a static configuration is used to produce a flexible behaviour that exhibits a complex but coherent musical personality in response to diverse musical situations. This is shown in Chapters 3, 5, and 6.

2.2.5 *VMMAS*

Another system, the *Virtual Musical MultiAgent System* (*VMMAS*), though not explicitly intended for free improvisation, is intended for other improvisational contexts (Wulfhorst et al. 2003). Although the initial implementation was trained to improvise accompaniment to popular music, Wulfhorst, et al. indicate that jazz performance is a potential future application. Like Collins (2006), the authors refer to modelling affective states with their system, initially focused on the simple binary of "happy" and "sad", which they link to rhythmic pulse. However, they indicate a more elaborate set of characteristics including "cautious", "flexible", and "persuasive", ascribed to simple parameter-related behaviours. For example, "cautious" agent behaviour would indicate that it plays only when it has a high "confidence degree" of detected input parameters.

In contrast to those of Collins' simulation, however, the agents of *VMMAS* do not communicate with each other directly, but rather, through a "blackboard". A blackboard architecture, commonly used in multi-agent systems, is characterised by a central point of coordination through which all inter-agent communication takes place (see Corkill 1991). When a human interacts with a collection of *VMMAS* "musician agents", any initiated change to the ensemble's tempo, meter, or tonality is first sent to the blackboard, before it can be addressed by the computer agents. Specific states internal to each agent determine whether and how new data will alter its output behaviour.

Although Wulfhorst et al. (2003) have designed *VMMAS* for real-time performance with human players, in their words, "to reduce the distance between musicians and computers", some significant shortcomings in their initial prototype were revealed in their study. The authors discovered that in the human musician's playing, "abrupt tempo changes should be avoided, because they can lead agents to wrong interpretations of the new tempo", and that "the harmonic changes must be evident in the musician's execution" (p. 590). Finally, they state that "a library of previously configured agents would be desirable, since it can be hard to configure every agent". These issues indicate that each agent's flexibility in adaptation, in particular, concerning responsiveness to real-time unstructured input, is significantly limited. This limitation is further addressed below.

2.3 Cybernetics

The constraints on the adaptive capabilities of *VMMAS* can be described with W. Ross Ashby's (1958) notion of 'requisite variety'. Ashby (1958) analyses the mutually regulatory effects of interacting subsystems in terms of a *requisite variety* of responses to the variety of conditions that demand a response. In other words, if a scale of values must be kept within certain limits to achieve a necessary equilibrium, the forces that might upset the balance must be matched with the appropriate (i.e., requisite variety of) counterbalancing forces.

For example, if blood pressure drops too low or gets too high, an organism will not survive, so various regulatory subsystems respond dynamically to maintain values within the necessary limits, a notion he refers to as 'homeostasis'. Ashby points out that biological systems with the implicit goal of survival are only one example of this phenomenon, and one could make the same analysis of an economic system, a chemical system, and so on, varying the goal accordingly. *VMMAS* did not have the requisite variety of responses to effectively sustain a collaborative performance with a human expert; the human performers needed to limit their own typical musical behaviour, so as to not overwhelm the limitations of the system.

A collaborative performance can be understood as emergent from the interaction between the human and computer agents, who act as coupled subsystems. This view of the general relationship between coupled subsystems and emergent phenomena is a key insight of cybernetics. While Norbert Wiener was viewed as the foundational figure of cybernetics in the United States, Ashby belonged to its foundational group in Britain, along with W. Grey Walter (Pickering 2010). It is interesting to note that Walter drew an analogy to collective jazz improvisation in his description of the coupled subsystems of the human brain:

> We know that within the brain, a great many electric processes can be identified, each with its own quite limited domain, some apparently independent, others interacting with one another. We must accept that in the EEG we are dealing essentially with a symphonic orchestral composition, but one in which the performers may move about a little, and may follow the conductor or indulge in improvisation — more like a jazz combination than a solemn philharmonic assembly. (Walter 1960, p. 17)

A similar view of coupled subsystems, in which each is capable of "temporary independence", is expressed in Ashby's *Design for a Brain* (1960 [1952]). Overall, cybernetic theory, in contrast to the formal-symbolic approach of classical AI, presents an alternative theoretical entry point for designing an artificial musical improvisor, and for investigating collaborative improvisation.

While cybernetic theory has been generally displaced by more current theories, it should be noted that Andy Clark, a contemporary philosopher of mind mentioned in the previous chapter, and Hillel Chiel and Randall Beer, leading researchers at the intersection of cognitive science, neuroscience, and robotics, have also drawn analogies between collective jazz improvisation and the tightly coupled, mutually regulatory subsystems of brains, bodies, and environments. The following quotations, which closely echo the cybernetic theories of Walter and Ashby, contain these analogies to improvisation, and are thus worth reproducing here at length. The Clark citation in the previous chapter is given here with additional context:

> In the case of biological brains and local environments it would indeed be perverse [...] to pretend that we do not confront distinct components. In cases where the target behaviour involves continuous reciprocal causation between the components [...] we do not, I concede, confront a single undifferentiated system. But the target phenomenon is an emergent property of the coupling of the two (perfectly real) components, and should not be "assigned" to either alone. [...] The players in a jazz trio, when improvising, are immersed in just such a web of causal complexity. Each member's playing is continually responsive to the others' and at the same time exerts its own modulatory force. (Clark 1997, pp. 164–165)

Clark's description supports the idea that in analysing a collaborative improvisation, while we can identify the roles of individual participants, there is an important sense in which the joint performance must be considered as a continuous whole, rather than as decomposable into discrete components or units of time.

Along similar lines, Chiel and Beer (1997) note that:

> The nervous system is often seen as the conductor of the body, choosing the program for the players and directing exactly how they play. The results reviewed above suggest a different metaphor: The nervous system is one of a group of players engaged in jazz improvisation, and the final result emerges from the continued give and take between them. In other words, adaptive behaviour is the result of the continuous interaction between the nervous system, the body and the environment, each of which have rich, complicated,

> highly structured dynamics. The role of the nervous system is not so much to direct or to program behaviour as to shape it and evoke the appropriate patterns of dynamics from the entire coupled system. As a consequence, one cannot assign credit for adaptive behaviour to any one piece of this coupled system. (p. 555)

Their notion that the "the final result emerges from the continued give and take between [players]" is precisely the way in which the collaborative human–computer performances with *Odessa* is analysed in subsequent chapters.

2.3.1 *CAIRA*

Jonas Braasch (2011) recognises the relevance of cybernetics to the study of free jazz and freely improvised music. Some of his insights appear to carry over into his work on *CAIRA*, a "creative artificially-intuitive and reasoning agent" that serves as an "improvisor and conductor" for musical free improvisation (Braasch et al. 2012). In particular, in his cybernetic analysis of collective free improvisation, he points out the importance of detecting significance in perception, and attempts to address this problem with extensive sonic analysis operations in *CAIRA*. This approach is in line with Pressing's (1988) notion that, in musical improvisation, "attentional emphasis [...] given to a particular [musical] component means that it will guide the generation of subsequent events" (p. 162).

Braasch also acknowledges that the process of detecting significance is subjective, and culturally and historically relative, that is, that there is no universal criterion for significance. In collective free improvisation, rather than a singular perspective spanning participants such as that implied by a score for notated music, multiple individual perspectives interact and affect future action. As Smith and Dean (1997, p. xiv) point out, during collective free improvisation, "a multiplicity of semiotic frames can be continually merging and disrupting" among the participants.[2] Given this diversity of perspectives within collective free improvisation, Braasch notes the centrality of the complex sonic information taken in by subjective listening, as opposed to a more universal mediation of sound as it is captured by conventional musical notation.

Thus, a performance of freely improvised music cannot be effectively reduced to a transcription. General support for this view can be found in music cognition research, which demonstrates a related phenomenon in the perceptual mechanism of cue abstraction. Deliège et al. (1996), in studies on listener perception of composed (notated) music,

[2]For a closely related point concerning the performance of composed works, see Levinson 1987.

identify two primary types of perceived musical cues: those that can be confirmed by consulting the musical notation — "'objective' cues (themes, registral usages, etc.)" — and, in contrast, "'subjective' cues, which have psycho-dynamic functions (impressions, for example, of development, or of commencement) which may be experienced differently from one listener to another and are not necessarily identifiable in the score" (p. 124).

As introduced in Chapter 1, Bailey (1993 [1980]) makes a similar point that extends this objective–subjective dichotomy into what could be viewed as a general quantitative–qualitative divide concerning different understandings of improvisation:

> Most improvisation has scant regard for the niceties of the tempered scale, or for exactly uniform divisions of the 'bar' or beat. [...] Even when man's senses are supplemented by such devices as the oscillator and the frequency analyser the result is only a more exact picture of the irrelevancies. It still has nothing to say about the forces behind the music making. (p. 15)

More recently, Sansom's (1997) doctoral research has further pursued the idea of examining aspects of improvisation that resist traditional notation and other forms of quantification. By investigating the social psychology of freely improvised music, Sansom (1997) focuses on general interconnections between musical structure and social dynamics, highlighting the ways in which nominally extramusical "information" — aspects of musical interaction not amenable to quantitative measurement — serves to shape the underlying processes of collaboratively improvised music, as a performance unfolds in real time. As Sansom describes his work, he uses a qualitative psychological and phenomenological approach to perform an "analysis of music's experientially defined dynamics of interaction and process" (p. 1).

In this respect, Sansom's approach relates to that of Sudnow (2001) (also referred to in the previous chapter), but, rather than an auto-ethnographic view of developing improvisational competence, as Sudnow presents, Sansom (1997) conducts an experimental investigation of the emergent collaborative dynamics of duet improvisation from the standpoint of an outside observer. For both Sansom and Sudnow, subjective human experience is regarded as the central object of research into improvisation, a premise supported by the research of Deliège et al. (1996) and Bailey (1993 [1980]); this perspective is also adopted for the research on *Odessa*.

Sansom's (1997) empirically derived model of free improvisation appears to support Braasch's (2011) theoretical cybernetic analysis of free improvisation. But the cybernetics-inspired design of *CAIRA* differs significantly from the early cybernetic approaches of Ashby and Walter. The design for *CAIRA* is exceptionally complex, using long-term audio signal evaluations to determine the "emotional force" of a musical section; a "'dictionary' of reference gestures" for interpreting musical gestures, with potential matches rated by probability-based "confidence" values; and a simulation of cognition using "logic-based reasoning" with an extensive ontology that includes concepts such as "musician" and "solo" (Braasch et al. 2012).

In contrast to such design complexity, Ashby (1960 [1952]) and Walter (1950, 1953) focused on the idea that a wide range of complex behaviours could emerge from a small collection of simple subsystems, tightly coupled but capable of temporary independence. On this basis, Walter built autonomous robotic tortoises (also called turtles), which, despite being far off from human-level intelligence, exhibited basic lifelike behaviour through their interactions with each other, with humans, and with their physical environment. As shown in the next chapter, the *Odessa* design follows a similar approach, favouring parsimony over complexity.

2.3.2 From cybernetics to Subsumption

The cybernetic view of the mind and of agent–environment interaction were at odds with the classical formal-symbolic, information processing view of the mind described above. But classical AI, in the vein of Newell and Simon discussed in the previous chapter, remained the predominant approach to robotics for several decades from the 1950s onward. At the end of the 1980s, however, Rodney Brooks noticed a contrast between the shortcomings of then-current AI and the decades-prior success of Walter's robots (see Brooks 2002, pp. 17–31).

The shortcomings of classical AI were, in some respect, similar to the limitations of *VMMAS* — early robots such as SRI's *Shakey* failed to live up to the demands of real-world interaction. Even an apparently simple robotic task like moving through an area while avoiding objects required software that could construct elaborate internal computational models of the external world, as well as hardware that could handle extensive calculation for planning and execution. Due to resource constraints, many such experiments were merely virtual; when they were physical, the environments needed to be almost completely stripped down — for example, a room with a white background, populated with simple solid-coloured geometric objects. Even with such reductive setups and tremendous processing power, this approach generally failed to do an effective job (Brooks 1999).

In response to these failures observed in the field, Brooks (1999) began to pursue an alternative approach, building insect-like robots and other biologically analogous models.

For Brooks, like Walter before him, the basic premise of his robots is that a small number of simple components, when interconnected in particular ways, can result in non-trivial complex behaviours. These behaviours can only be observed over extended periods of time, and are thus best described at higher levels of abstraction.

Brooks formalised this design premise into a specification he termed the 'subsumption architecture' (also referred to simply as Subsumption), that describes the construction of interactive "layers" of "competing behaviours", from which global behaviours emerge. The mechanically embodied robots built using Subsumption were rapidly prototyped with "real world" empirical testing, in contrast to the white walls and carefully placed solid-coloured geometric objects of previous robotics experiments. When the prototyping was complete, the mobile Subsumption robots could navigate complex, dynamic, unstructured environments, such as ordinary offices with real furniture and people moving about constantly (see, e.g., Brooks 1999; Mataric 1990; Connell 1989).

It can be noted that Brooks has expressed a narrow view of musical communication by suggesting that improvisation must require visual, rather than purely sonic interaction (see Lewis 2007). In more abstract terms, however, the previous chapter indicated a potential conceptual link between the idea of *musical improvisation* and Connell's use of the concept of *improvisation* to refer to the capabilities of a *non-musical* (mobile) Subsumption robot. Thus, despite Brooks' domain-specific view of music, a more widely applicable notion of improvisation is relevant to AI research. As Lewis (2007) points out, a renewed definition of improvisation should be capable of "usefully reconnecting supposed purely musical questions with their analogues in similar issues surrounding the practice of everyday life itself" (p. 116). Thus considered, research into computer improvisation — musical or otherwise — fits with the more general aim of robotics research to build "autonomous artificial cognitive systems that are to pursue their goals successfully in real-world environments that cannot be fully anticipated, that are not fully known and that change continuously, including other agents" (Müller 2012, p. 1).

The *Odessa* design in some sense follows from the connection between Subsumption and improvisation. It is particularly striking the way in which the improvised navigation of a crowded office by a mobile robot may resemble the abstract navigation of a musical environment in free improvisation. This resemblance is effectively revealed with a conception of the musical environment in social terms; however, as considered below, there are other ways of applying Subsumption to music that more closely relate to traditional music theory.

2.4 Other musical systems related to Subsumption

2.4.1 *Reactive Accompanist*

Joanna Bryson (1995) was the first to develop a musical agent using Subsumption (and, more generally, the first to apply Subsumption to software development). Her system, the *Reactive Accompanist* thus relates to *Odessa*, in so far as both are Subsumption agents for music. Moreover, although she does not refer to improvisation as the agent competence she seeks to evaluate, her description does, however, imply the evaluation of an improvisational competence. She refers to the evaluation of a "folk" approach to music, which, in her account, corresponds to the way in which real human instrumentalists (folk musicians) can skilfully elaborate a real-time accompaniment to an unknown melody, without the benefit of a score (p. 6). In addition, her research uses a qualitative evaluation methodology based on human assessment, which is, in this respect, similar to the research on *Odessa*, although the methodological details differ significantly (see Chapter 4).

Despite these general similarities, it is considerably difficult to directly compare the two systems. This is due to significant differences in both the nature of the musical competencies being modelled by the systems, and specific implementation details that relate in part to design decisions, and in part to changes in the state of technology since the time her system was developed. I will highlight these differences here.

The aim of Bryson's system is "to derive chord structure from a melody in real time", which "emulates the human competence of providing chord accompaniment to unfamiliar music" (p. 20). She clarifies that her system should produce a "harmonious accompaniment to the melody", although she acknowledges that "just what is 'considered harmonious' is subjective" (p. 20). Musicians and lay persons were used for her system evaluation to "judge whether the chord structure of the piece 'sounds reasonable' " (p. 20; see also pp. 70–72).

The first point of divergence relates to the notion of "real time". As she states, "due to difficulties with signal processing of the input, the programs are not actually in real time, but the processing they do assumes that they are" (p. 87). There are several issues to identify here, beginning with the fact that she is faced with the disadvantage that no "off-the-shelf" real-time Fourier transformer was available to her, a considerable drawback that stands in stark contrast to today. However, she contradicts the point that real-time processing is assumed by her system programming, stating that, if real-time processing were available, "there would be some redesign involved in the main functions of the robot programs, because in a real-time system one would not sample the *next* input, one would sample the *current* input" (p. 82, original emphasis). Of course, she is pointing out a logical implication, but it underscores the difficulty in comparing her system with one that performs in real time, such as *Odessa*.

In addition, in contrast to *Odessa*, her system is constructed from several neural networks that must be trained in advance. This has the implication that, "as what it hears becomes further from its trained input, its performance gradually degrades" (p. 80). *Odessa* does not use neural networks, which makes it more parsimonious, and it does not require any advanced training, which makes it more flexible with respect to performance context.

The next point relates to the *Reactive Accompanist*'s modelled capacity "to derive chord structure from a melody". Most of the individual competencies that work together to achieve this aim necessarily affect one another in a reciprocal fashion. Nevertheless, her strategy can be thought of as a "bottom-up" approach: identify note boundaries in the input stream, identify the pitch of each note, relate the pitch to a tonal centre, match the tonal centre to a chord (stored in advance), and monitor the tonal centre for a break that would require a new chord. In short, the system "rapidly stabilises to the chord which is the primary key of the melody" (p. 61), and "the rhythmic perception competences [...] offer reasonable locations to break off and look for a new chord", which "results in much more key-compatible chords being produced as output" (p. 59).

The tonal logic of Bryson's system calls for a reduction of pitch frequency to pitch class (non-octave specific diatonic note names), and the system depends upon classifying input in terms of tonal key, which it matches to "a priori" (stored) chordal information. *Odessa*, on the other hand, uses original (received) input frequencies within the system, although these may be mapped to (e.g.) notes on the piano keyboard at the output stage (as is the case with the implementation used in the present study). Moreover, *Odessa* does not match input frequencies to a tonal key, as such matching is not a strict requirement for free improvisation. In addition, in contrast to Bryson's system, *Odessa* does not look for a regular beat to inform the timing of its output, as free improvisation does not strictly require a regular beat. Moreover, while the chordal accompaniment of a melody is modelled as a "following" behaviour in Bryson's system (the accompanist follows the lead), *Odessa* may also lead rather than follow, or engage in a construction of lines that parallel, but do not match, the human performer.

Finally, although I have indicated some underlying similarity between the evaluation methodologies of these systems, Bryson placed considerably less emphasis on her evaluation. By her own declaration, "the evaluations were carried out fairly informally" (p. 70). She used two musicians and herself (also a musician) to evaluate symbolically represented system output (a melody annotated with conventional chord symbols). She also used two musicians and herself, plus three lay persons, to listen to and evaluate the system's auditory output (on the basis of recordings, in lieu of real time performance). As shown in Chapters 4–6, the evaluation of *Odessa* was given significant emphasis in the present investigation.

2.4.2 *BeatBender*

Another documented musical system that describes itself as using Subsumption is Aaron Levisohn's and Philippe Pasquier's (2008) *BeatBender*. While Bryson's system is primarily focused on harmony, *BeatBender* is focused on rhythm and does not explicitly take pitch into account (the percussive samples they use could be said to have a quasi-pitched characteristic). In a broad sense, their system serves as a musical exploration of how simple interactions between simple rules can result in complex output, which is a general characteristic of Subsumption systems. However, although the only available technical description is insufficient to make a precise determination, it seems their system would be more aptly described as a generative looping multichannel sequencer, rather than as a Subsumption system.

Their system is presented as a multi-agent system in which each agent controls a dedicated audio channel. All activated channels are mixed together equally to form an audio output stream. Each channel is dedicated to a single looping audio segment; across channels, all segments have an equal duration and all are synchronised. For each iteration, a set of conditional rules determines where one or more sound events will occur in each channel, at various positions within the segment. This results in continuously changing rhythmic patterns.

In sharp distinction from Subsumption (and, by extension, *Odessa*), all agents in their system share common environment variables, which suggests some similarity to a blackboard architecture (see Corkill 1991; see also Brooks 1999, p. 97, further discussed in Appendix A.2.1). And, significantly, no agent receives audio input from outside of the system. The sound made audible to human observers is strictly a result of human-configurable options and agent interactions within a purely virtual environment.

2.5 Psychology

The idea that a continuously changing, unpredictable real-world environment also includes other agents is important. Building on this idea, a theoretical relationship can be established between ecological psychology, which deals with agent–environment dynamics, and more traditional empirical psychological research, such as developmental psychology, that address the issue of how we identify other agents within our environments. Theoretical insights from both fields are discussed below, in relation to the research on *Odessa*.

2.5.1 Ecological psychology

The relationship between physical and musical environments is explored in cognitive terms by Eric Clarke's (2005a) "ecological approach to the perception of musical meaning". Clarke draws on an ecological and enactive theory of perception originally developed by James Gibson (1966, 1977, 1979). Ecological psychology fundamentally concerns itself with the tight coupling of perception and action, or, more broadly, of agent and environment, in the context of adaptive behaviour; Gibson (1977) is also the source of the well-established notion of affordances. Simply put, an affordance is a feature of an environment that presents a possible action to an agent, relative to the agent's capabilities. Using Gibson's theory of affordances, Clarke examines the complex interpersonal and environmental interrelations that underpin collaborative musical improvisation (see Clarke 2005a, pp. 152–154). His discussion refers to embodied and situated cognitive negotiation in both physical and musical terms, thereby demonstrating an ecological link between perception, action, and meaning.

Clarke's (2005a) view of action-oriented listening in improvisation ties in with Braasch (2011), Sansom (1997), and Sudnow (2001), in their consideration of the relation between performing musical improvisation and listening for subjectively identified significance, which may vary from moment to moment. All of their views have in common the idea shared by ecological psychology and phenomenological philosophy, cited in the previous chapter, that "how we directly pick up significance and improve our sensitivity to relevance depends on our responding to what is significant for us" in a given context (Dreyfus 2008, p. 361). It should be noted that contemporary philosopher of mind Clark, cited above, also finds parallels between the conceptual frameworks of ecological psychology and phenomenological philosophy in the work of Gibson and Merleau-Ponty (Clark 1997, chap. 8.8), as well as between Gibson's ecological psychology and Brooks' robotics research (Clark 1997, chap. 2.6).

In the previous chapter, it was pointed out that a connection can be identified between Brooks' research, in particular, his development of Subsumption, and ethology research based in ecological psychology. The kinds of agent–environment interactions that are the focus of both were also shown to be related to cybernetics and more recent cybernetics-influenced research. The ecological perspective also relates to cybernetics through the notion of requisite variety, because the diversity of musical possibilities afforded to a performer through listening makes possible a diversity of musical responses.

Thus, performer interactions can result in a complex collaborative dynamic, or, in other words, a collaborative performance can emerge from the performer interactions. In their empirical study of a string quartet, Davidson and Good (2002) state that "music-making [...] provides players with many affordances, each player taking the 'meaning' that is specified at any particular point in the performance as a 'means' for further specifying what is afforded by [...] the mutually constituted 'musical product' of the joint activity of all the players" (p. 200). Though not explicitly referring to improvisation, the idea of a mutually constituted outcome of joint activity relates to the notion of an emergent collaborative performance that has been discussed throughout this chapter.

2.5.2 Intentional agency

From the perspective of a performer of improvised music, the significant aspects of sound that may guide musical responses are enhanced by the presence of affordance-rich material (material that suggests many possible responses to an attentive listener). When co-performers are mutually responsive, a tight coupling with "continuous reciprocal causation" can result in an emergent collaborative dynamic. However, from the same perspective, one could describe the relation between a performer and a musical instrument, as in the case of a solo performance.

A further distinction is required to understand the role of another performer. If a sonic-musical environment suggests the presence of a co-performer, the sensitivity to significance or relevance can be importantly altered. In particular, it can be altered through the inference that the co-performing agent can potentially behave as an equal participant in a shared activity, like the members of the string quartet referred to above. This depends not only on the inference of the presence of another agent, but also that the agent's exhibited behaviour may be interpreted as intentional, that is, as if its behaviour is aimed at producing an intended result. On this basis, one can interact with an intentional agent as a collaborative co-performer in ways that are not afforded by interaction with a typical musical instrument.

Current research in psychology suggests that an interplay of adaptation and resistance ultimately leads to the attribution of intentionality, which can be brought to bear on the design of human-interactive artificial agents. Psychological research into the attribution of intentionality can be traced back to an early empirical study of adults, which found that they were prone to interpret certain movements of animated geometric shapes as the actions of persons (Heider and Simmel 1944). Although philosophers of mind have long been concerned with theoretically unresolvable questions concerning *intrinsic* intentionality (see Dennett 1987), current empirical psychology research on the *attribution* of intentionality has played an important role in contemporary cognitive modelling (e.g., Baldwin and Baird 2001) and biomedical research (e.g., Castelli et al. 2002).[3]

Studies in developmental psychology indicate that the human predisposition to attribute intentional agency to both humans and non-humans appears to be present from infancy. Poulin-Dubois and Shultz (1988) chart childhood developmental stages over the first three years of life, from the initial ability to identify agency (distinguishing animate from inanimate objects) to the informed attribution of intentionality, by inference of goal-directed behaviour. Csibra (2008) found that infants ascribed goal-directed behaviour even to artificially animated physical objects, if the objects were secretly manipulated to suggest teleological actions, such as avoiding obstacles to reach a marked point. Király et al. (2003, p. 767) identify the source of an infant's interpretation of a teleological action: "If the abstract cues of goal-directedness are present, even very young infants are able to attribute goals to the actions of a wide range of entities even if these are unfamiliar objects lacking human features". The "even if" is particularly relevant to the research on *Odessa*, as it supports the idea that an interactive improvising music system could be effective in engaging a human co-performer without necessarily being human-like.

Concerning the notion of resistance, a study of adults by Barrett and Johnson (2003) suggests that intentionality can be perceived in an interactive context even when a self-propelled (non-human) object has no positive goals, only negative ones, such as avoiding a pitfall. Their test subjects used language normally reserved for humans and animals to describe the behaviour of artificially animated physical objects that appeared to exhibit resistance to direct control in the course of an interaction; when there was no resistance, they did not use such language. The authors of the study link the results of their controlled experiment to the anecdotal experience of frustration arising in interactions with artefacts such as computers or vehicles that "refuse" to cooperate. In other words, in an interactive context, if an artificial agent is too passive and does exactly what is expected of it, this may negate any sense of its apparent intentionality. This suggests that for an agent to remain apparently intentional during direct interaction, it must exhibit a degree of resistance in addition to adapting its behaviour to the environment.

Müller's (2011) examination of general HCI also points to a connection between resistance and perceived intentionality. In this context, he finds that computing research can

[3]Dennett (1987) argues from a philosophical standpoint that attributed intentionality is the only form of intentionality needed for a fully developed account of the mind.

benefit from what is referred to as 'thick' description in social anthropology, meaning not just a description of what people do, but also of how they experience what they do, in relation to their general experience. This emphasis on experience is also reflected in broader ethnographic trends within HCI research (see Dourish 2006). In light of the above-described studies of improvised music focused on experience, and the focus on experience in both psychological research on intentional agency, and ethnographic research on HCI, it is clear that the related research on *Odessa* must focus on experience as well.

2.6 Methodological background

The focus on experience in the present investigation has a number of methodological implications, which are considered below.

2.6.1 Early AI and qualitative human judgement

Regarding the difficulty in evaluating Subsumption systems in general, Brooks (1999, pp. 74–75) emphasises the contrast between traditional AI and Subsumption when he states that, for Subsumption systems, although *a priori* knowledge can be "incorporated into a robot, [...] it must be non-specific to the particular location that the robot will be tested in". He continues by stating that Subsumption robots can thus "seem to have much more limited abilities" than traditional AI systems, which, unlike the former, are never confronted with an "uncertain" and "coarsely described world" (p. 75). Here, he makes the point that for traditional AI, evaluation is simplified with an idealised design and experimental setup. One example of this is when a robot has been programmed to identify a predesignated object, and then tested with precisely that object, which has been carefully placed so as to remove any potential interference with the identification.

Odessa uses Subsumption principles to accommodate a wide variety of input. On this basis, *Odessa* faces the Subsumption-related problem described by Brooks, in that one can not evaluate the simple success or failure of solutions to narrowly-specified problems. Instead, one must consider the relationship between a set of simple mechanisms (based on general *a priori* assumptions), and the general capabilities and versatility of the system.

For *Odessa*, as Brooks' (1999, p. 73) points out with respect to his mobile robots, this suggests that evaluation must not be held to the criteria of once traditional academic robotics and AI; these typically consider only the kinds of results that were presented early on in these disciplines, primarily in the mid-1960s through the mid-1970s, especially regarding the research conducted in the Artificial Intelligence Center at the Stanford Research Institute (later renamed SRI International) (Nilsson 1984). According to Brooks, his design approach can not be adequately evaluated by using criteria developed for earlier, fundamentally different systems and their corresponding experimental conditions. While Brooks does not make any explicit recommendations, the relevance of qualitative evaluation to Bryson's (1995) musical Subsumption system was discussed in section 2.4.1.

Even in the strongly mathematically oriented work of Norbert Wiener (1948), founder of cybernetics, the suggestion is made that qualitative observation and interaction must ultimately determine the fitness-for-purpose of an AI design, in particular, for a chess-playing machine. Wiener (1948) notes that one could theoretically imagine a chess-playing machine that could evaluate every possible allowable option, and thus play an optimum game every time.[4] But, Wiener continues, such a machine would not exemplify the kind of performance exhibited even by human experts. Yet, as he observes, a chess-playing machine that simply follows the rules of the game would be trivial, as the rules alone do not address the merit of play. Even the most current chess playing systems fall between these two extremes: they do not merely play an arbitrary game by following the rules without any apparent strategy, while, at the same time, they use heuristic methods to narrow their search for possible strategic moves, rather than considering all possible games.

As Wiener (1948, p. 165) argues, for a satisfying human–computer interaction in playing chess, an appropriately constructed machine should not only follow the rules of chess, but play it at a level "not so manifestly bad as to be ridiculous". In other words, "it may attain a pretty fair level of accomplishment". He summarises the capacity of such a machine by saying "it would probably win over a stupid or careless chess player, and would almost certainly lose to a careful player of any considerable degree of proficiency". (Today, advanced systems can defeat even leading human chess experts, but typical consumer chess software allows the playing level to be adjusted for an evenly-matched game, similar to what Wiener imagines.)

Wiener's example invoking subjectively evaluated computer performance relates closely to the case of *Odessa*, for which a Subsumption design allows it to interact with expert players who may exercise their expert playing capacities. As in Wiener's example, these experts are, in the course of an interaction with the system, expected to encounter behaviours that are not necessarily equivalent to those of expert-level human players.

[4] As Haugeland (1985) points out, a practical implementation of this idea must use a 'generate and test' approach, which could not actually evaluate all possible games on a human time scale.

Yet, the system performance is also not expected to degrade into what may be considered incompetence.

Another mathematically oriented thinker, Alan Turing, also turns to qualitative evaluation in his theoretical examination of interactive computer behaviour. In his paper on what has come to be referred to as the 'Turing test', Turing (1950) puts forth the idea that a computer's ability to hold a conversation is demonstrative of its intelligence. His proposed experiment involving a 'question and answer'-based conversation with a computer is suggested as an interactive, qualitative measurement that would in principle determine if a computer has met its design goal, namely, to produce intelligent verbal responses. The assessment of the computer's responses is based solely on human judgement, as people are generally assumed to be experts in everyday conversation, capable of discussing general and perhaps specialised topics. Indeed, Turing even considers the idea that a computer's *incorrect* answers to a human-posed question may be a greater sign of human-like intelligence than if it were solely to produce correct answers.

The above examples from Wiener and Turing highlight the relevance of dynamic interactivity, as Brooks does. In a similar role to the human chess player or interlocutor in the above examples, those interacting with *Odessa* are experts in performing collaborative free improvisation. Their judgements of the system, on the basis of their own interactions with it, are an effective qualitative measure of *Odessa*'s capacity to engage in the practice of collaborative free improvisation. This notion of evaluation differs from the types of evaluation that measure computer output on the basis of formal rules for determining correctness, which in some cases may be appropriate, depending on the design goals of a system.

This section has shown that questions of evaluating AI computer system performance have, since the beginning of AI, dealt with qualitative information, such as whether a series of conversational responses demonstrates intelligence (Turing 1950), or whether a chess position is "strong" or "weak" (Wiener 1948).[5] A similar issue has also arisen with Bryson's (1995) musical Subsumption system, where a qualitative evaluation of her system's harmonic accompaniment to a human performer was found to be necessary. This was due to the fact that the computer-generated harmonies her system produced could not be regarded as merely correct or incorrect; rather, the harmonies were among multiple formally correct possibilities that were evaluated by experts in harmony as being more or less preferable. More generally, Brooks (1999) has called for an evaluation of Subsumption robots that recognises the dynamics of interaction between the robots and their unsimplified deployment environments.

[5]Dreyfus and Dreyfus (2005) describe "the lack of precise and situation-free definitions" of "situational aspects of [chess] positions [such] as a weakened king's side or a strong pawn structure".

The next section will further explore how, generally speaking, deployment context significantly bears on system interaction and performance evaluation.

2.6.2 Experimental context and participant perspective

The above section considers the evaluation of the type of interactive systems for which there is a close relationship between certain aspects of a situation and apparently intelligent behaviour, such that the interplay between agent and environment is constant. If a situated agent is considered in complete abstraction from the situation, intelligent behaviour becomes more difficult to identify, or, as Brooks (1999, p. 68) puts it, for situated agents, "the boundary between computation and the world [is] harder to draw".

Irrespective of methodology, when seeking to experimentally evaluate a system, the context in which the system will perform and be evaluated in must be considered. For example, a physical mobile robot for exploring rugged terrain is different from a disembodied expert system for use in a hospital, and both are different from one solely designed to analyse linguistic structures. As Brooks (1999, p. 128) states, "a medical expert system, or an analogy program cannot climb real mountains. It is clear that their domain of expertise is somewhat more limited, and that their designers were careful to pick out a well circumscribed domain in which to work".[6] Having a well circumscribed domain for a system is not only relevant to system design, but also relates to the evaluation of the system performance.

An interactive system is designed for one or more deployment contexts, e.g., a musical collaboration, a game of chess, or a verbal conversation, such that the evaluation must take the design goal into account. For example, a quantitative measurement of speed would serve to evaluate a robot designed to give a *fast* answer to a question, but would not necessarily evaluate a robot designed to give a *good* answer. In this respect, the design goal in part determines how the evaluation should be carried out.

For the present experimental investigation, the participants performed a series of freely improvised duets with *Odessa* and reported on their experience (details of this process are presented in Chapter 4). They began with the foreknowledge that they would be interacting musically with a computer agent. Arguably, some of the agent's deficiencies are unconsciously made up for in the process of interpreting its dynamic behaviour (Collins 1990). However, an agent's sufficient performance of a complex practice can be regarded as evidence that it exhibits at least a basic capacity to engage in the practice.

[6] In this context, Brooks (1999, p. 128) does, however, argue that physically grounded systems will eventually "solve" the "whole problem" of intelligence, an argument which is not considered here, as it is beyond the scope of the present research.

By analogy, a study examining the listener-perceived accuracy of pianists performing familiar composed works found a high rate of perceptual tolerance for technical deviations in performances that did not violate the musical context of the score; pronounced violations of the score were more readily detected (Repp 1996).

Although faithfulness to a score is not directly relevant here, it is important to emphasise that unreliable perception and tolerant interpretation cannot make up for all deficiencies. A further study could present participants with an experimental control using input-independent output, to compare and assess to what degree performance with any output might be interpreted as collaborative. For the present study, *Odessa* uses neither entirely fixed nor entirely random output, but rather, it is to some degree responsive. Rather than testing if and how responsivity is detected using a non-responsive control, the study instead assesses whether and to what extent the given responsivity is interpreted as being part of a broader, mutually collaborative improvisational activity.

Generally speaking, the relationship between perception and interpretation is complex. Neurobiological research into perception indicates that when perceptual data is ambiguous, prior experience guides one to a (potentially incorrect) 'best guess' to resolve the ambiguity, for the purpose of a rapid interaction with the environment (Ernst and Bülthoff 2004). This points to the fact that there may be differences among individuals in how they independently resolve ambiguities in a shared environment. Music cognition research by Deliège (see, e.g., Deliège et al. 1996; Deliège 2001, 2007) shows evidence of a related phenomenon in the perceptual mechanism of cue abstraction.

Deliège's theory of cue abstraction was first mentioned earlier in this chapter, in a discussion of detecting significance in improvised music. That discussion also included a citation from Pressing (1988, p. 162) stating that, in musical improvisation, "attentional emphasis [...] given to a particular [musical] component means that it will guide the generation of subsequent events", as well as one from Smith and Dean (1997, p. xiv) stating that, during collective free improvisation, "a multiplicity of semiotic frames can be continually merging and disrupting" among the participants. When these citations were introduced, I hinted at their relation to Deliège's theory of cue abstraction, but at that point, I did not yet explain the underlying connection to Deliège. Rather, in that context, I indicated their relation to cybernetic theory, in contrast to traditional AI, where traditional AI was characterised by its treatment of intelligence as consisting of symbolic representation and "problem solving by search" (Newell and Simon 1976, p. 120).

One interesting aspect of Deliège's (2007) theory expressly pertains to the idea of search, and further highlights the divide between an ecological theory of improvisation and the traditional AI paradigm of system design — a divide that in part motivated my use of Subsumption for *Odessa*. Deliège (2007) acknowledges that not all music is built according to "theme and variations", which she illustrates largely with examples from modern compositions, but her point can be extended to include improvisation. She identifies two distinct general cognitive processes of musical listening: one that occurs when listening to variations on a theme (or other related cases of listening for pre-identified elements), and another that addresses more open-ended musical listening. In particular, when listening to music based on "theme and variations",

> the psychological situation of the listener is [...] mainly geared to a kind of premeditated search — which is therefore *explicit* — for comparisons while listening. But [...] such a premeditated search is only possible if the listener has been able to hear the reference elements beforehand, as in the case of pieces built on the 'theme and variations' type of composition, and this is not necessarily the case. (Deliège 2007, p. 28)

When music does not overtly follow such a pattern, "the listener does not strive *consciously* to detect iterations; rather, he or she *implicitly* submits to their impact because of accumulation" (Deliège 2007, p. 27, original emphasis).

The accumulation of iterations Deliège refers to relates to the abstraction of *cues*: "A person who is *listening with attention* will seize cues, that is, brief but meaningful and significant structures, which stand out from the sound background" (Deliège 2007, p. 13, original emphasis). As Deliège (2001, p. 238) describes, "the cue [...] provides us with a basic point of reference for the comparisons between musical structures that occur throughout the listening process". These points of reference and comparisons differ among individual listeners on the basis of their attentional focus and memory, both of which are affected by experience. This sense of subjective listening relates to individuals' everyday experience and formal learning, as well as to the frequently varying psychological and physiological conditions pertaining to attention and memory recall.

As Wiggins (2007, p. 332) notes, Deliège's framework for understanding musical similarity on the basis of cues is such that "more individualistic views of [musical] similarity may be expressed", which "is important because of the lack of agreement between individual respondents about musical similarity". It is in this respect that the points about improvisation made by Pressing (1988) and Smith and Dean (1997) can been seen to interrelate to those of Deliège. Namely, for each individual, as Deliège (2001, p. 238) points out, "different cues that are abstracted during listening [act] as waymarkers or milestones", which "gives rise to the notion of a *mental line*". This subjectively and implicitly detected musical structure is a key point of reference for the improvisor that will thus "guide the generation of subsequent events" (Pressing 1988, p. 162).

Meanwhile, as individuals recognise or abstract cues during improvisation, they may at times seize the same cues as each other, and at times different ones, which can lead to "semiotic frames [...] merging and disrupting" (Smith and Dean 1997, p. xiv). In other words, subjective perceptions of musical structure among individuals for a given source may alternately overlap and diverge in key respects. This relates to the musical application of the ecological theory of affordances in so far as participants in a collective improvisation may pick out different structures to respond to not only consciously, but also at the most basic level of perceptual activity.[7]

As *Odessa* is not a biomimetic model, it does not implement an internal mechanism of cue abstraction. However, this mechanism is nevertheless related to the present investigation in that the cue abstraction by the human participants facilitates the treatment of *Odessa*'s output as affording musical responses on different time scales. While *Odessa* is deliberately limited in that it may react only to immediate input, the long-term accumulation of perceptual cues in the experience of the human co-performer allows for both immediate reactions and reactions to affordances that may emerge more gradually.

2.6.2.1 Analysing interactivity

Though not directly related to music, Joseph Weizenbaum's (1966, 1967) ELIZA program, designed to verbally interact with a human participant, is a widely known example in the history of interactive systems. Its design differs fundamentally from that of *Odessa*: ELIZA is based on linguistic pattern matching and rule-based analysis and generation, whereas *Odessa* uses none of those methods. While the design details of the former are not relevant here, the interactive context is.

Weizenbaum points out that *framing* the system performance for participants, by influencing their expectations, significantly affects the interpretation and evaluation of the system behaviour. In the case of ELIZA, the human participant is led to expect a specific type of conversation; in the case of *Odessa*, the participant is led to expect

[7]The issue discussed here concerning different, simultaneously possible perceptions of structure, and responses that follow from the perceived structure, also relates to research on the *Copycat* system, which uses AI to study creativity in human cognition (Mitchell and Hofstadter 1990; Hofstadter 1994). Hofstadter (1994) describes what is referred to as "Problem 5", where the system must find a mapping between "abc" and "mrrjjj". If the relationship between them is conceived of by the program on the basis of letters, by extrapolation, the related "abd" might in turn be mapped to "mrrkkk" (the last letter of each sequence is replaced by the letter of the alphabet that follows it). However, if the relationship is conceived of in terms of group-lengths, the discerned 1-2-3 pattern can lead to the mapping of "abd" to "mrrjjjj" (the 'j' repeats four times for 'd', extrapolated from its three repetitions for 'c'). The confluence of the account of the physical domain of musical perception and action described above and that of the conceptual domain of analogy described here suggests a fundamental cognitive mechanism that underpins both. See Hofstadter (1995), especially Chapter 5 for more on *Copycat*, and Preface 9 for a discussion of issues pertaining to the experimental evaluation of models that resonates with the discussion in the present chapter.

a collaborative free improvisation. In both cases, participants interpret the systems' behaviour within a specified, delimited context.

In the original papers describing ELIZA (Weizenbaum 1966, 1967), a conversation is presented as something that can be intuitively recognised, without the requirement of a specific experiment or method of evaluation. Example conversations with ELIZA (actually, with its DOCTOR variant) are given as evidence that the system is fit to hold a conversation. However, rather than any variety of conversation, the verbal interaction is framed in a certain way for the human interlocutor, namely, as one would talk to a Rogerian psychiatrist, meaning that a speaker is encouraged to continue speaking in response to the psychiatrist's prompts. Weizenbaum gives an example of an exchange with ELIZA in which a human participant mentions a boat, eliciting the system response, "tell me about boats"; such responses, although simplistic, serve to direct the conversation and keep it going. According to Weizenbaum, this allows for the system to converse without recourse to any knowledge of the world (similarly, Subsumption systems, by definition, have no explicitly represented knowledge of the world).

Weizenbaum uses the example of ELIZA to raise questions about the notion of machine understanding. With respect to the computer's internal mechanisms, it clearly does not use language in the same way that humans do, and generally lacks a technical apparatus that would allow it to have even the possibility of understanding. But it is important to emphasise that, from the perspective of the participant, the performative act of conversing can take place, analogously to how humans might converse. In other words, the system provides the participant with a certain kind of interaction, and this interaction can potentially become, for the participant at least, what he terms a "plausible" conversation (presumably meaning plausibly realistic or human-like).

As Weizenbaum (1967) states:

> No understanding is possible in the absence of an established global context. To be sure, strangers do meet, converse, and immediately understand one another (or at least believe they do). But they operate in a shared culture — provided partially by the very language they speak — and, under any but the most trivial circumstances, engage in a kind of hunting behaviour which has as its object the creation of a contextual framework. Conversation flows smoothly only after these preliminaries are completed. (pp. 475–76)

His account can be thought of as a model of conversation. By using a certain design for ELIZA, and setting up a global context for interaction via instructions to the human participant, the empirical evaluation of ELIZA's performance can contribute to the validation of this model. As Weizenbaum's research indicates, in addition to the global context of interaction, the on-going interaction between the participant and the program provides further context, which can in turn facilitate a plausible conversation.

To summarise this section thus far, Weizenbaum's work on ELIZA presents a sociolinguistic model of conversation and interactive software that lends support to his model. Both the interactive context and the software design are relevant to establishing his model. The software is designed to fulfil a human role in a typically human interaction, assuming that a particular context is given by expectations that are instilled in the human participant.

Weizenbaum makes clear in his research that the computer program lacks human-like understanding and, in some cases, he even discloses this to those interacting with his system. Yet, from the perspective of humans interacting with ELIZA, or those observing such interactions, a conversation does in fact take place. Thus, even if the computer uses language in a substantially different manner than what is typical of human linguistic competence, the human participant's interactive behaviour may provide insight into sociolinguistic interaction, for example, on the roles of inference and interpretation.

Building upon this approach to theory, design, and experimental evaluation, for *Odessa*, rather than investigate the more general problem of *machine improvisation*, which does not suggest a clear standard of measurement, the present research attempts to reveal insights about collaborative improvisational interaction between humans. As with ELIZA, it does so by taking a typically human–human interaction, and substituting one human with a computer that can fulfil the requirements for a collaborative improvisation to take place. Human participants interacting with *Odessa* can report on their experience, and their interactions with the system can also be externally observed.

2.6.3 Methods applied to the study

As the musical interaction experience was the central consideration of the study, it was decided to use a qualitative approach with traits in common with ethnography and some psychology and human factors research. Some specific considerations for this decision are presented here. This is followed by an explanation of the general methodology of the study, which informs the specific methodology discussed in Chapter 4.

Research on improvisation has revealed phenomenological (Sudnow 2001) and psychosocial (Sansom 1997) dynamics that cannot be accounted for with analyses of western notation. Musicological studies of the bodily (Givan 2009) and sociocultural (Monson 1994) forces that shape improvisation further underscore the fact that improvised music is irreducible to traditional musical formalisms. These findings reinforce broader accounts of human-world relationships, both specific to music (Clarke 2005a) and more generally (Dreyfus 1992 [1972]).

Davidson (2004) offers a well-balanced account of the variety of quantitative and qualitative empirical research methods that address music as social behaviour. With a more narrow focus in the present study on musical interaction experience, it was critical to "ensure that the assumptions built into measurement procedures correspond to the structure of the phenomena being measured" (Hammersley 2010, p. 15). As Clarke (2004, p. 91) points out, a quantitative approach "entirely misses the social dimension of performance", including the "interactions between performers", at least in so far as some of the social interaction is experiential rather than externally manifested.

While Dourish (2006) provides insight into the tensions between ethnographic and HCI research agendas, Ball and Ormerod (2000) suggest a middle ground between strict disciplinary approaches:

> One might see the ethnographic features of situatedness, participant autonomy, personalization, reflexivity and independence as implying precisely the converse of a list of features that define the methods of experimental cognitive psychology, namely reductionism, experimental manipulation, impersonalisation, objectivity and hypothesis-driven data sampling. However, [...] the methods of experimental cognitive psychology are not always appropriate for exploring contextually rich multi-agent domains such as design, and [...] the methods of pure ethnography lack feasibility in the human factors domain. (p. 152)

Along similar lines to this position, a compromise was sought for this research, in which experimental economy and controls were important to the study design to ensure a justifiable comparison across data gathered from multiple cases; at the same time, an open-minded stance was taken during the analysis, to allow unanticipated insights to emerge from the collected data.

Overall, this approach shares commonalities with the experimental method, but also has a lot in common with ethnography, in that it values a richness of detail and does not disregard information that falls outside of preconceived notions about the phenomena under investigation (see also Flyvbjerg 2006 on the relation of scientific enquiry to qualitative case study research). In more general terms, "ethnography typically involves participant-observation in a process of tacking between culturally immersed practice and distanced reflection" (Williams and Irani 2010, p. 2730). For the present study, not only the researcher, but also the participants were called upon to engage in this "tacking" process in order to effectively judge the system performance.

Semi-structured interviews were used to explore the participant perspectives. Wengraf (2001) writes of how semi-structured interviews typically require more preparation and more analysis than fully structured interviews. This is because in a semi-structured interview, one must closely pay attention to the answers, understand them in the context of the interview and the study as a whole, and engage in a more conversational role with the participant. In contrast, in a fully structured interview, one proceeds through a series of questions without any regard for the content of the answers during the interview.

Wengraf (2001, p. 25) states that, as an interviewer in the semi-structured mode, "I both respond to what has been said (and not said) so far in the conversation, but also act in the present in anticipation of possible futures of the conversation, which I wish to move towards (or to avoid)". This type of conversation is "much more *artful*" than is suggested by turn-by-turn or question-and-answer models; sometimes, for example, an "interviewer goes back to [a] 'dropped hint' at a much later point and gets a different dimension [of the topic] that was present but not observed at the time" (p. 41). His approach is consistent with my experience posing follow-up questions designed in the moment to elicit responses relevant to the research, while remaining neutral so as to allow both positive and critical feedback.

The method of analysis I have used largely follows the tenets of the "analysis method framework", also known simply as 'Framework'. This approach to analysis, " now widely used by qualitative researchers" (Ritchie and Lewis 2003, p. 220), is related to other qualitative research practices, including interpretive phenomenological analysis (IPA), grounded theory, and discourse analysis. But as Ritchie and Lewis (2003, p. 201) state of qualitative analysis in general, while there are many different traditions and methods, with different notions of what should comprise the main assumptions and focus of analysis, "distinctions are not always clear cut". As they point out, and as is the case here, boundaries are often crossed between traditions, or within individual studies.

In Framework,

> [a] thematic framework is used to classify and organise data according to key themes, concepts and emergent categories. As such, each study has a distinct thematic framework comprising a series of main themes, subdivided by a succession of related subtopics. These evolve and are refined through familiarisation with the raw data and cross-sectional labelling. (Ritchie and Lewis 2003, p. 213)

The result is a "cross-sectional analysis based on interpretations of meaning" that ultimately aims at "making sense of the findings through the production of descriptive and explanatory accounts". The authors describe the way in which, using these methods, there is (analytical) "movement both up and down" the stages of the analysis.[8] In this way, "categories are refined, dimensions clarified, and explanations are developed". This takes place while constantly reviewing the data "to check assumptions or to identify underlying factors," with the goal of sensibly representing the original material.

2.7 Summary and conclusions

Following Brooks' (1999) notion of an architecture comprised of interactive behavioural "layers", it would seem that a system decomposed into an active "agency" layer, a reactive "adaptation" layer, and an additional "resistance" layer could implement an interactive computer music system that is perceived as an intentional agent. If it is to be perceived as engaging in a collaborative improvisational dynamic with a human co-performer, it must also provide an affordance-rich variety of material, thereby enabling human co-performers to engage in their typical collaborative behaviour. Subsumption offers rapid, flexible responsiveness to unstructured input, in virtue of its lack of a complex central representational model of the world, and its lack of a central locus of control from which pre-formulated decisions are executed. These characteristics should facilitate the performance of musical free improvisation, which can be regarded as unstructured input that requires rapid and flexible responses.

The central aim of this chapter has been to present literature relevant to the following chapters. Alongside more general considerations, it examined how other systems relate to the research on *Odessa*:

- The *VMMAS* (Wulfhorst et al. 2003) ultimately lacks a requisite variety of affordances, leading to limited possibilities for motivated musical responses. This has served as a general point of reflection concerning the selection of a Subsumption design, which allows for complex dynamic responses to unstructured input on the basis of a small collection of interconnected simple mechanisms, with no internal representation.

- *Voyager* (Lewis 1999, 2000), a successful system that embodies intentional agency and offers affordance-rich output, does so on the basis of the complexity of its internal representation, which is tuned to produce the desired system behaviour.

[8]This is depicted in Chapter 4, Figure 5.

Examined here is the extent to which the proposed system, *Odessa*, could succeed without a similarly complex internal representation.

- *Swarm Music* and *Swarm Granulator* (see, e.g., Blackwell and Young 2004a) model the natural phenomenon of animal swarming with equivalent behaviours among swarm particles or agents; the designers find a strong aesthetic connection between swarming and emergent phenomena in collective free improvisation. In contrast, Subsumption systems model a collection of diverse behaviours, so it remains to be shown that *Odessa*'s Subsumption design can lead to similarly musically coherent emergent phenomena.

- *CAIRA* (Braasch et al. 2012) implements a complex internal model as well, although in this case, the system is based on theories of human cognition with respect to internal cognitive mechanisms; it thus also serves as a testing ground for such theories. *Odessa*, in contrast, aims to produce perceptual cues that are used by humans in interpreting behaviour and subjective musical meaning, without a complex internal model, and without necessarily replicating human cognitive mechanisms within the agent.

- *Cypher* (Rowe 1992, 1993) is also based on a model of human cognition, but, like the Subsumption approach, it regards intelligence as a collection of competing behaviours. In contrast to a Subsumption design, however, *Cypher*'s behaviours are organised in such a way that internal representations are first constructed, then evaluated by a deliberation mechanism, and then executed as musical output. (This contrast is discussed in further detail in the next chapter.)

- The *Free Improvisation Simulation* (Collins 2006) uses a small collection of simple algorithmic agents capable of rapid responsiveness and resulting emergent complexity. In this respect, it has some similarity to a Subsumption design. However, unlike a Subsumption design, the agent behaviours are constrained at design-time by the configuration of individual parameters associated with personality traits.

- The *Reactive Accompanist* (Bryson 1995) was the first musical system to use Subsumption, and it appears to be the only other musical system to conform to this architecture. However, Bryson did not aim to model free improvisational behaviour. Primarily due to this difference, the specific design of *Odessa* is largely unrelated to that of Bryson's system. Moreover, due to historical technological limitations, Bryson's system did not actually run in real time, thus making *Odessa* the first interactive musical Subsumption system to do so.

Building on the research described above, I proposed that an effective artificial agent for collaborative human–computer free improvisation could be implemented using Subsumption. I have presented a significant number of motivations for such a design, and evidence that such a musical system has not been previously implemented. This points to a potential contribution that would address a significant gap in existing research.

I have also supported the idea as to how perceptual cues for intentional agency could emerge from the interaction of behavioural layers in the collaborative improvisational agent, and in the overall interaction of the system with a human co-performer. I have indicated that this could be achieved without a complex internal representation or central locus of decision-making and execution, but rather, with a small collection of interconnected low-level simple mechanisms. This could in principle result in the perception of a musical identity or personality that is not specified at design time.

As opposed to computer music systems evaluated according to traditional musicological criteria, especially using composition derived studies of notation, current empirical research on musical improvisation, including free improvisation, points to the centrality of experience and collaboration between participants. Thus, an appropriate evaluation of the proposed system must be based on a theoretical understanding of the role of experience in collaborative improvisation.

One important connection between the system design and the experimental evaluation can already be noted here. Frequently, research in computational cognitive science centres around complex models of narrowly defined phenomena. This approach is also commonly found in computer music research, such as the psychoacoustic or cognitive musicological modelling of listener expectation (e.g., Milne et al. 2011; Pearce and Wiggins 2012). Such phenomena may be viewed as related to what Beer (1996, p. 422) describes as "minimally cognitive behaviour", which "is meant to connote the simplest behaviour that raises cognitively interesting issues". As he states,

> on the one hand, the capabilities and behaviour of the model agents that we study must be rich and sophisticated enough to be cognitively interesting, so that they raise the sorts of issues that we would like to explore. [...] On the other hand, these model agents must be simple enough to be computationally and analytically tractable, so that we have some hope of evolving and analysing them using techniques that are at most an incremental step beyond what is currently known to be feasible. (Beer 1996, p. 422)

Yet, while this in some respects pertains to the design of *Odessa*, the goal of the present research is not to establish a biomimetic model, but rather to investigate cognition through the study of how human participants respond to a system designed to provide certain cues for interaction. In this respect, the proposed model is perhaps more similar to the earlier chapters of Braitenberg (1984), where simple models facilitate the attribution of emotional states; the later chapters in that work, however, move more in the direction of Beer (1996). In short, *Odessa* is not designed to be internally similar to humans, but the evaluation nevertheless aims to investigate human cognitive mechanisms, as described further in the remaining chapters.

The next chapter will explain how the ideas discussed in this chapter are implemented in the system design. It will provide a detailed account of Subsumption in general, and also provide the specifications for the implementation of *Odessa*.

Chapter 3

Building *Odessa*

Odessa is programmed using Subsumption, developed by Rodney Brooks (and his students) since 1985, beginning at the MIT Artificial Intelligence Laboratory (Brooks 1999). Subsumption, an approach to building physically embodied robots, was developed, in part, as a response to purely software-based simulations and, in part, as a response to resource-intensive physical robots (Brooks 1999). This chapter explains the way in which Subsumption has been adapted for and applied to *Odessa*, and also provides a detailed account of the system.[1]

A basic explanation of some key terms will be helpful to the following discussion. The concepts presented here are elaborated further in the in-depth description of *Odessa* below; prior to that, these concepts may be referenced in passing, where a cursory understanding will suffice to understand the points being made. Subsumption robots, traditionally physically embodied mobile robots, consist of a collection of special-purpose *modules*. A module — an independent physical or machine-simulated processor — is an augmented finite state machine (ASFM) with multiple input and output connectors (Brooks 1999, p. 13). An ASFM is a conventional finite state machine (a set of possible states and conditions for state transitions) that is augmented with registers (to which data is written) and timers (alarm clocks). These registers and timers can conditionally trigger state transitions (Brooks 1999, pp. 40–41). Although a module is sometimes referred to as an agent, particularly when describing a Subsumption robot as a multi-agent system, I will use the term *agent* only to refer to the entire system (i.e., the full set of interconnected modules).

Subsumption robots are built by combining individual modules into networks called *layers*, which are also interconnected. A layer typically implements a particular behaviour, and the behaviour implemented by the lowest layer of control is called the level 0 (or

[1] Additional background on the implementation is given in Appendix A.

zeroth level) competence (Brooks 1999, p. 16). For a mobile robot, a level 0 competence (behaviour) might be to avoid obstacles. A higher level behaviour, e.g., a level 1 behaviour, might be to wander around a room. The interaction of behavioural layers such as these comprise the general ability of a Subsumption robot, which may be, for example, to navigate around an unstructured, unmodelled dynamic environment (e.g., a busy office). The robust handling of unstructured, unmodelled dynamic environments is a fundamental feature of agents built using Subsumption.

3.1 Architectural principles: Using Subsumption for music

Overall, earlier robotics and AI implementations often had slow response times, not only due to slow processors, but partially due to elaborate ontologies and reasoning components; this meant that "systems that were built could not respond to a dynamic real world" (Brooks 1999, p. 68). One motivation for developing Subsumption was to facilitate the construction of systems that would be robust and flexible enough to deal with real-world conditions. Prior to Subsumption, systems were developed typically in either purely software-based environments or ideally simplified physical laboratory conditions. Robots built under either of these conditions did not translate well to real-world deployment scenarios (Brooks 1999, pp. 73–74). In contrast, the Subsumption approach favours early deployment in the environment for which systems are intended. Thus, there are initially low expectations of a robot's performance, but its capabilities are continually improved upon with on-going minor adjustments. This iterative process continues until a robot is robust enough to meet the challenges that it typically encounters in its unsimplified target environment.

Similar concerns and motivations to those behind Subsumption were shared in the development of *Odessa*, despite the fact that it is not a mobile robot, but rather, a musical agent. *Odessa* was designed to collaboratively improvise with human experts performing in an unsimplified manner. Its development followed the Subsumption strategy of avoiding idealised models of input, while also avoiding an excessively complex technical apparatus that could not be deployed early and rapidly improved upon. Some general themes common to Subsumption research that pertain to *Odessa* are described in the following subsections.

3.1.1 Real versus virtual input/output

Although *Odessa* is not physically embodied with actuators, its sonic output is physically manifested via a loudspeaker or electromechanical instrument. This "real world" sound constitutes the output with which human performers interact. For human instrumentalists interacting with *Odessa*, instrument sound is captured via a transducer (microphone or pickup), with an analogue-to-digital conversion via an off-the-shelf sound card. In this way, real physical sound is the basis for (human–computer) agent interaction.

This situation is opposed to that of using simulated software environments, where virtual agents interact solely on the basis of data that is representative of physical sound, despite the absence of physical-sonic input into the system (e.g., Unemi and Bisig 2004; Beyls 2007; Levisohn and Pasquier 2008). In such systems, even when sound is made audible to a human observer, agent interactions take place within a virtual environment, and interactions are affected only by internal system data.

3.1.2 Computational resources

Odessa runs on a standard personal computer and thus has significantly more processing speed and throughput than typical embedded microprocessor-based Subsumption robots. Generally speaking, for a real-time system, a given computational process must be completed within a certain timespan (e.g., 100 milliseconds) to be relevant to a specific task. To complete this (hypothetical) process in time, a (hypothetical) embedded microprocessor may take too long (e.g., 1000 ms), whereas a (hypothetical) standard personal computer may potentially complete it in far less than the required time (e.g., 10 ms). The point of this example is to illustrate that, for real-time systems being developed on standard personal computers today, developers may potentially achieve their desired results without necessarily using the most efficient computational approach. Continuing with the above example, if a 10 ms process completion were in principle attainable, but completion takes 90 ms in the implementation, the suboptimal implementation may have no noticeable impact on system performance.

The issue of computational resources is significant because Subsumption robots were originally designed to more effectively use limited computational resources to achieve robust real-time system performance. These resource constraints may appear less relevant to research today that uses personal computers. However, while *Odessa* indeed draws upon available processing resources for intensive tasks such as real-time audio analysis, it also continues in the Subsumption spirit. *Odessa*'s ability to respond quickly to changes in a dynamic real world is achieved by using computationally lightweight mechanisms and short connections (i.e., few interventions) between system input and output (the robotics equivalent of sensors and actuators). A detailed account of these characteristics is given below in section 3.2.

3.1.3 Unstructured input and models

The Subsumption approach to handling input and output, which can also be thought of as an approach to agent–environment interaction, is highly suggestive of the cognitive engagement by performers of free improvisation. During free improvisation, performers exhibit tight coupling between listening and playing (input and output). They also have a robust, flexible approach to dealing with unpredictable changes in the environment (Clarke 2005a; see also Sudnow 2001).

One of the motivations for developing Subsumption was to "respond quickly to changes in the world" (Brooks 1999, p. 68). This responsiveness is, in part, achieved by a means of accommodating unstructured input, as opposed to expecting input to conform to an internal model. This is addressed in this subsection. (It should be noted that responsiveness is also achieved is by using short connections between input and output, discussed in section 3.2.)

Brooks' (1999) work has shown that his robots function effectively without the use of internal models, that is, without ideal formalisations of the outside world, which tend to limit responsiveness. *Odessa* follows the Subsumption approach of eschewing models, using Brooks' insight that "the world is its own best model" (Brooks 1999, pp. 115, 128). This aspect of Subsumption is significant in relation to previous approaches to modelling improvisation.

Certain forms of improvisatory music, although not solely based on musical formalisms (see, e.g., Bailey 1993 [1980]), have proved amenable, at least in simplified form, to formalised musical description, such as rules for fitting a melody to a chord progression (e.g., Biles 1994). While there has been some work on formal models of free improvisation, these have typically relied upon non-musical formalisations, such as dynamical systems models (see, e.g., Blackwell and Young 2004a, for an application to system design; and, e.g., Borgo and Goguen 2005, for a theoretical musicological analysis). Although a non-learning system such as *Odessa* is limited in certain musical respects by its lack of internal models, the present research concerns its collaborative role in human–computer interactive free improvisation. It is hypothesised that *Odessa*'s ability to collaborate with experts in this domain is not compromised by these formal musical limitations.[2]

3.2 System design and implementation

This section is organised as follows:

- Input and output (3.2.1)
    - Note stream formation (3.2.1.1)
    - Note stream decomposition; contrast to similar systems (3.2.1.2)
- Interactivity (3.2.2)
    - Interaction model (3.2.2.1)
    - Interactive behaviour; background; layer decomposition (3.2.2.2)
    - Subsumption layer interaction (3.2.2.3)
- Detailed description and explanation of modules (3.2.3)
- On the development of the initial prototype (3.2.4)

3.2.1 Input and output

3.2.1.1 Note stream formation

Implementing a mechanism for continuous musical output, that is, for a continuous *note stream*, poses a challenge when seeking to adhere to Subsumption principles. Some background will help to clarify this challenge. The Subsumption approach was, in part, developed as an alternative to what is known as the *sense–model–plan–act* conception of robotics. This conception led to robots that used a linear series of information processing modules that consisted of: *sensing* input data from the world (via sensors); identifying sensed objects by fitting input data to a preconceived *model* with a rigid ontology; using a collection of rules to reason about those objects and formulate an explicit *plan* about what to do next; and, finally, performing some *action* using actuators to produce an effect in the world. For robots built in this way, "an iteration through the cycle could

[2]See John Stevens (1985) for one approach to free improvisation that is neutral with respect to formal musical abilities. In the present discussion related to Brooks' artificial insects, it is perhaps fitting to mention that "in the wake of John Stevens and the Spontaneous Music Ensemble, certain strands of English improvised music were known, half-disparagingly, as insect music" (Toop 2004, pp. 238–239).

often take 15 minutes or more" (Brooks 1999, p. 63). Alternatively, Subsumption robots are designed to limit the processing time between sensors and actuators, so that input (perception) affects output (action) in a continuous series of short cycles of alternation between them. This rapid perception–action cycle, inspired by biological systems, can be understood as forming a tight coupling between agent and environment. With this tight coupling, "the boundary between computation and the world [is] harder to draw, as [Subsumption] systems [rely] heavily on the dynamics of their interactions with the world to produce their results" (Brooks 1999, p. 68).

The notion of a tight coupling between agent and environment via a perception–action cycle has been theorised in a recent cognitive theory of improvisation, based on principles of ecological psychology (Clarke 2005a). The recognition of commonalities between this ecological theory of improvisation and Subsumption contributed to a core hypothesis of this research, namely, that a Subsumption agent could be effective for collaborative free improvisation. However, by definition, Subsumption agents lack high-level representation, so there is no straightforward way to achieve the coordinated integration of short input–output cycles into a long-term global representation. From a traditional perspective, this would seem to pose a difficulty for the construction of a continuous musical output stream that may include musical phrases, rests, and textures.

Continuous stream formation in *Odessa*, consistent with the Subsumption principle of short cycles, is depicted in Figure 1. It is achieved as follows: discrete monophonic note streams, in a continuous series, are *passively* integrated into a *continuous polyphonic* note stream. The stream formed by this process is *continuous* and *polyphonic*, as multiple segments are spawned before other segments (audible sequences of notes) have terminated. This results in overlap between the discrete segments, which provides continuity and also serves to form chords and complex rhythms. The integration is *passive* because the monophonic note streams are spawned without any explicit coordination, other than their successive delivery to the sound producing mechanism (synthesised or acoustic).

Although no notation is used in the actual system, Figure 1 depicts an approximation of the note stream formation process as an imagined transcription of a musical section (the bottommost staff). The upper four staves show four note sequences independently generated in complete cycles (their relation to input is described further below). In the actual system, such sequences are neither globally quantised nor synchronised to one another in a musical sense (although they are subject to the same processor clock, as described in Appendix A.2.2).

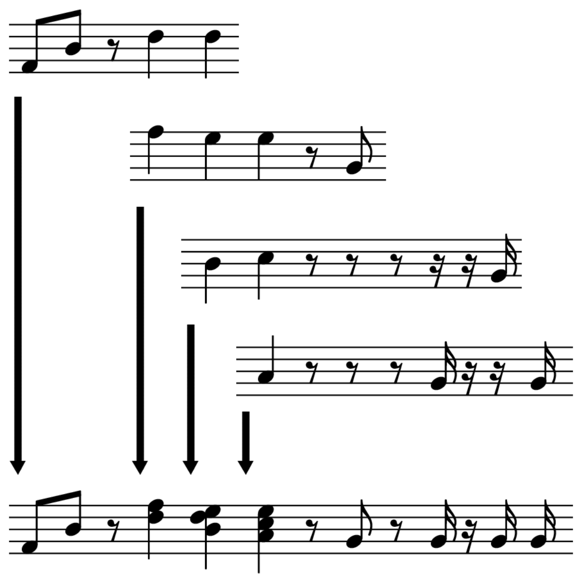


Figure 1: Approximation of streaming output formation (arrows indicate merging). No traditional notation is used in the system, nor is there any note-level synchronisation or quantisation.

3.2.1.2 Note stream decomposition

In the literature on interactive music systems, a common design abstraction is a functional decomposition into listening, analysis, and performance (e.g., Rowe 1992; Lewis 1999; Wulfhorst et al. 2003; Blackwell and Young 2004a; Assayag et al. 2006; Hsu 2010, etc.). *Odessa* is decomposed differently, following Subsumption principles that result in a distribution of components without a central locus of representation or control (Brooks 1999). This contrast is depicted in Figure 2.

Separate subsystems are used for pitch, loudness, and timing, respectively, in both input and output. As would be expected of a Subsumption system, *Odessa* uses no formal musical knowledge such as scales, tonal keys, motifs, etc., and also lacks any representational model.[3] In contrast, one or more of these means are used in the systems described in Chapter 2, such as representation of Western tonal harmony (e.g., Rowe 1992), stored motifs (e.g., Collins 2006), and representation of notes as a particle swarm (e.g., Blackwell and Young 2004a).

With *Odessa*, incoming sound to the soundcard (transduced via microphone or pickup) is analysed in separate, uniparametric dimensions: frequency for pitch approximation, amplitude for loudness approximation, and timing between the notes. These parameters

[3]The semitones used in the output are an artefact of the mapping of input frequencies to the nearest available frequencies produced by the output mechanism. The decision to use a piano (or synthesised equivalent) as the output mechanism is discussed in Chapter 4.

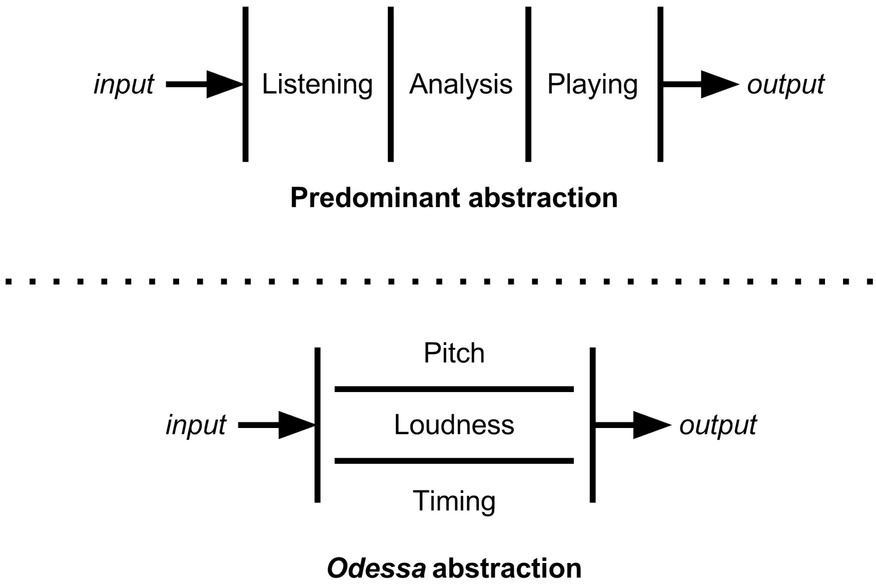


Figure 2: *Top:* Traditional horizontal decomposition with layers between input and output, as with sense–model–plan–act;
Bottom: Subsumption vertical decomposition in which each layer connects input to output (adapted from Brooks 1999, p. 67).

are concurrently analysed by dedicated modules (i.e., one for each). Output is formed through an integration of these separate parameter streams in a module that spawns short segments (sequences of notes), described above in section 3.2.1.1.

A similar approach to dealing with computer-based musical information was proposed by Conklin and Witten (1995), with their notion of "viewpoint decomposition". "Viewpoints" are independent abstractions for "expressing events in a sequence" in terms of a single parameter of a musical event's "internal structure" (e.g., pitches, intervals, durations). To form complete musical sequences, a variety of individual abstractions are recombined into "linked" viewpoints.

It is worth noting that this approach is not without precedent in the arts. Composer Milton Babbitt (1962) suggests a method of serial composition in which the pitch-based organization is complemented with a parallel independent organization of "time points", thus separating pitches from their temporal placement in theory before linking them in the final score. As early as 1939, filmmaker Sergei Eisenstein describes a technique found in the work of authors, directors, and actors that combines and juxtaposes "a few basic *partial representations*" such that an "integral image [...] *arises* [...] in the spectator's perception" (Eisenstein 1947, pp. 30–31, original emphasis).

Conklin and Witten's technique was specifically developed for probability-based analyses of a corpus in the service of generating new works similar to those in the training set. In contrast to this and related approaches such as Cope's (2005), *Odessa* does not use probabilistic analysis. Instead, it uses simple transformations of input from both externally- and internally-generated sources (this process is described in section 3.2.3).

3.2.2 Interactivity

3.2.2.1 Interaction model

A distinction between two common meanings of the word 'system' in software development has been pointed out by computer scientist Michael A. Jackson (2001, p. 11): There is the narrow sense of a computational system that is generally comprised of hardware with installed software; and, there is also a broader system that includes the narrow system, its deployment environment, and its users. The narrow system cannot be effectively designed without an understanding of the broader system. Although it may not be appropriate to think of an independent robot agent as having *users*, it is nevertheless the case that humans interacting with *Odessa* are part of a broader system. It is within this broader system that the collaborative interaction between artificial and human agents takes place.

It is thus relevant to discuss *Odessa*'s interaction model, in addition to its narrow system properties. As noted above, Subsumption agents, by design, "[rely] heavily on the dynamics of their interactions with the world to produce their results" (Brooks 1999, p. 68). And, as previously noted, Suchman (2007) points out that these "interactions with the world", for Brooksian mobile robots, are "understood primarily in physical terms", "evacuated of sociality" (p. 15). But for a *musical* Subsumption agent, the agent–environment interaction may indeed be social. This is especially the case for an agent that performs free improvisation, a practice that arguably consists of a fundamental psychosocial dynamic (Sansom 1997; see also Davidson 2004).

Odessa is designed to interact with human improvisors as an individual participant in a shared collaborative performance. This approach to the human–computer relationship differs from Pachet's, whose *Continuator* may appear to be related to *Odessa*; significantly, the *Continuator* is presented (e.g., in Pachet 2003) as a means to extend an individual's musical performance capacities (and is thus not considered here in depth). Pachet's system uses machine learning as a basis for its ability to musically interact, in contrast to *Odessa* and other systems such as Hsu's (2010). But while the latter's interaction abilities are tailored to specific instrumental techniques, *Odessa* is designed to function with a wide variety of instruments and players.

By responding to and introducing affordance-rich material into a collaborative context, *Odessa* adopts a model of interaction characteristic of musical free improvisation between humans. In terms of human–computer interaction, the nature of this model is encapsulated by Lewis' (1999) description of his free improvisation system, *Voyager*, described in Chapter 2. As he states of *Voyager*, "there is no built-in hierarchy of human leader / computer follower, no 'veto' buttons, pedals, or cues" (Lewis 1999, p. 104). This general approach to interaction design is shared by other systems with similar aims, such as those by Blackwell and Young (2004a) and Collins (2006), although implementations vary greatly (see Chapter 2). More generally, this interaction model is opposed to "game-theory models of social interaction that emphasise self-interest", and instead emphasises coordination, "interdependence", and "mutual control" (Young 2010, p. 97). The human and computer players function as tightly coupled subsystems, exerting a constant reciprocal influence on one another.

3.2.2.2 Interactive behaviour

Collaborative musical free improvisation is a form of interaction between distinct individuals who collectively negotiate the construction of musical pieces in real time, without anything agreed upon in advance (Bailey 1993 [1980], pp. 83ff). Thus, an artificial agent must sufficiently convey to a collaborative human co-performer that it is listening, responding, cooperating, adapting, and also that it is a distinct entity capable of making independent musical contributions. The collection of these and similar capabilities points to an agent's (apparent) *intentionality*, which, more generally, describes its ability to carry out actions in a certain way; namely, its actions suggest that it understands its relation to the environment. Research in psychology, discussed in Chapter 2, suggests that a combination of perceptual cues — perceived when observing and interacting with an agent — lead to the attribution of intentionality.

For *Odessa*, one goal of the design was to produce such cues to convey intentionality, in order for interactions with the system to reflect the character of collaborative free improvisation. A Subsumption agent, described as a "collection of competing behaviours" (Brooks 1999, p. 90), lends itself to the production of such cues when the agent's behaviours are organised as an interplay of adaptation and resistance, an idea based on insights from psychology that is described below. For the design of *Odessa*, the musical sense of adaptation has been interpreted as an adaptation to the musical behaviour of the human co-performer, while resistance has been interpreted as producing a divergence from the human behaviour, to potentially lead the collaboration in a different musical direction. Thus, *Adapt* and *Diverge* form distinct higher-level behaviours of the system, while the basic capacity to *Play* (initially described above in section 3.2.1 and further described below in section 3.2.3) forms the lowest-level behaviour (level 0 competence).

When designing *Odessa*, it was hypothesised that this behavioural decomposition into *Play*, *Adapt*, and *Diverge* would serve to produce cues that suggest intentionality. (The empirical experimental results of this hypothesis are described in Chapter 5.) The basis for this hypothesis is found in psychological research into the attribution of intentionality, as discussed in section 2.5.2.

3.2.2.3 Layer interaction

The *Play*, *Adapt*, and *Diverge* behaviours of *Odessa* are separated into Subsumption layers, as depicted in Figure 3. In Figure 3, using Brooks' convention, the circles marked 'i' indicate *inhibition* and those marked 's' indicate *suppression*. When data is *inhibited*, the data is blocked from transmission along the line of data flow between modules. Blocked data is simply discarded, as described earlier, since no recipient is dependent on receiving data. When data is *suppressed*, data flowing from one module is replaced by data from a different source module. Brooks (1999) describes this as 'masquerading', because the recipient does not 'know' the data is coming from an alternate source; that is, incoming data received by a module is not treated differently on the basis of what module is sending the data.

In the absence of external (sonic) input from a human co-performer, the *Play* layer generates an independent musical output stream (initially described above in section 3.2.1 and further described below in section 3.2.3). When external input is detected, the *Adapt* behaviour is activated, which results in the output stream adapting to the human co-performer's musical behaviour by using pitches, loudness, and timing derived from and closely related to the input source. This can give the human performer a sense of *Odessa* cooperating.

However, if this layer remains activated for an extended period, the behaviour could be perceived as too passive, thereby negating the sense that *Odessa* exhibits intentionality. Thus, when a timer expires in the *Adapt* layer after it is active for a certain period, the *Diverge* layer is activated. The initial duration of the timer is set to a restricted pseudorandom value that is typically between 5 and 15 seconds. This value is recalculated each time the timer is reset after expiry, so as to be irregular and unpredictable. An equivalent version of this timer is found in the *Play* and *Diverge* layers, to prevent them

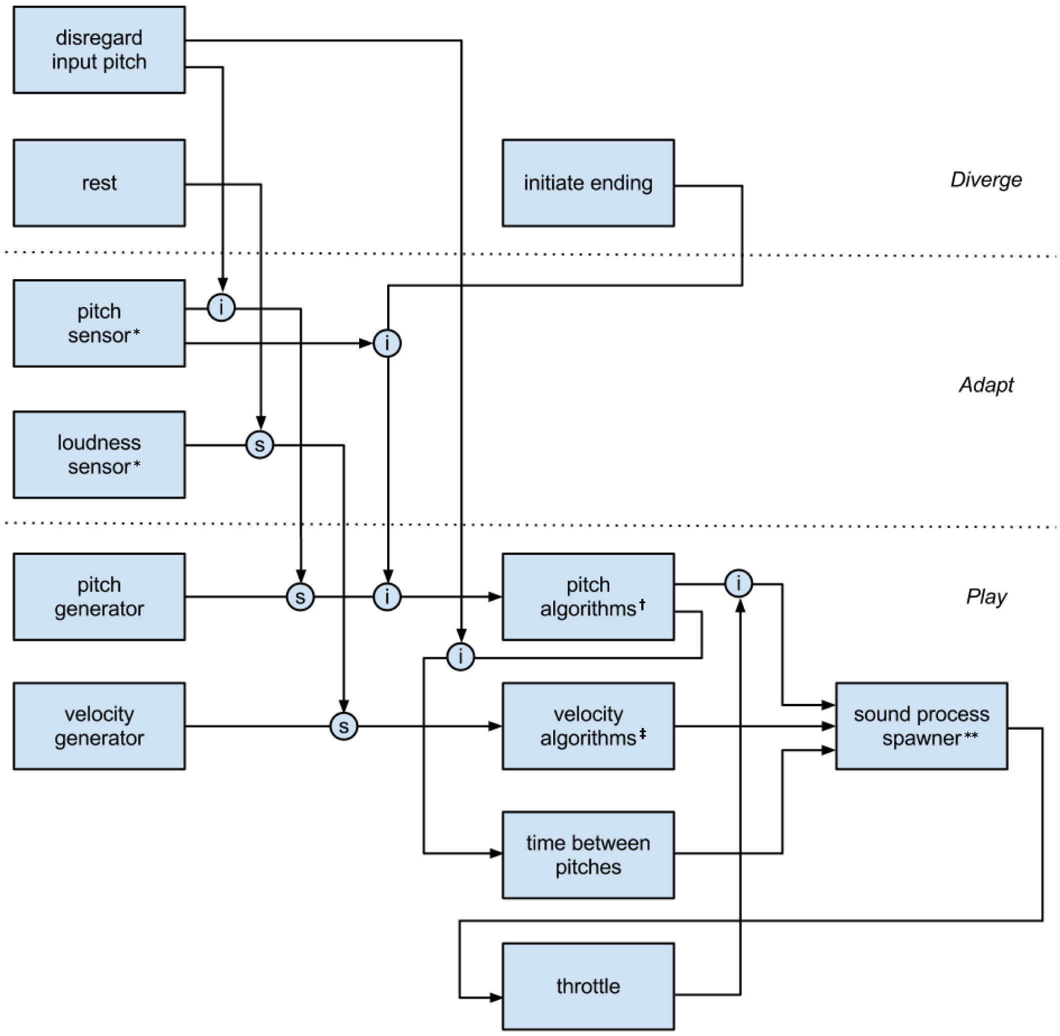


Figure 3: *Odessa* architecture. Modules are indicated by named boxes (see 3.2.3 for detailed module descriptions). Layers are separated by dotted lines. Solid lines indicate data flow in direction of arrow. Circles marked 'i' indicate *inhibition* and those marked 's' indicate *suppression* (see 3.2.2.3). *Receives external audio input. **Transmits external audio output. †Transforms input to output values by raising or lowering one semitone, or leaving them unaltered. ‡Translates input value into a collection of neighbouring output values.

from being active for too long. The *Play* timer range is also typically 5–15 seconds, and the *Diverge* layer uses different timers for each module (described in section 3.2.3.3).

The result of these timers is a dynamic interplay between layers. This interplay allows for the human co-performer to perceive the system's ability to react to input, and its ability to introduce different musical material. The human co-performer may not necessarily respond to such different musical material, but this is also the case in strictly human performances of collaborative free improvisation.

3.2.3 Individual modules

The potentially surprising simplicity of *Odessa*'s modules, when considered in the context of their practical roles in the system behaviour, form a key strength of this research.[4] One aim of developing an artificial agent for collaborative free improvisation using Subsumption is to demonstrate that complex interactive behaviour, subject to evaluation by experts, can emerge from interactions between simple modules, that is, modules that perform mathematically basic operations. Thus, at every instance where a more complex operation *within* a module could be substituted, I have opted instead to use a simple variant (presented in further detail below). The deliberate use of what might be considered "mathematically uninteresting" operations is significant because, for computer-generated music, it is well known that a mathematically interesting process can become a sonically interesting process when certain mappings between them are used (e.g., for constructing melodies, harmonies, rhythms, orchestrations, etc.; see, e.g., Xenakis 1992). For *Odessa*, if complex operations were used within the modules, the source of complex interactive musical output could not be exclusively attributed to the interactions between the modules.

Another point related to Brooks' comments above is that, when working with numerical data, several different mathematical representations of how that data is handled might, in principle, be accurate. Yet, when a simple numerical operation is used to achieve a practical aim in the system, it would be misleading to use a complex mathematical formalism to represent it. For example, consider a module U0 that generates a series of arbitrary integers to trigger notes on a piano, with the aim of producing a melodic line. In this case, one could potentially define the series of integers in mathematical terms as a *random walk*, which may be an accurate representation; moreover, such a

[4] Brooks (1999, p. 3) notes that although his original paper on Subsumption has become the most referenced paper he has written, at the time of its 1986 publication, it was "shocking" to senior roboticists, "because it argued for simplicity rather than for mathematical complexity of analysis and implementation". He adds that many people in the field "feel that their work is not complete if it does not have pages of equations, independently of whether those equations shed any light at all on the deep questions".

concept may be intrinsically interesting to mathematicians as a stochastic or Markov process, a self-avoiding walk, etc. (Barber and Ninham 1970). However, the important point for a Subsumption system is the basic mechanism and its relation to the aim of the system, subsystem, or specific module that implements the mechanism. Here, the initial description of U0, or perhaps a more detailed account of it, would be more relevant than a mathematical representation, because its role in the system behaviour is to facilitate a higher-level emergent behaviour, rather than to preserve any formal properties of low-level data.

The remainder of this section describes the mechanism and purpose of each module depicted in Figure 3, where the names in boxes correspond to module names. The descriptions are organised according to layers. It may be useful to consult Figure 3 for a system-level perspective of the interrelationships between modules.

3.2.3.1 Modules in the *Play* layer

- **Pitch generator.** Generates a pseudorandom number within the piano keyboard range (frequency translated to a MIDI value). It then pseudorandomly generates 5 pitches within a range of one octave below and one octave above the initial pitch. An array of these 5 pitches comprises the module output. When this module is active, the two-stage generation mechanism prevents too many octaves on the piano from being played simultaneously, which tends to suggest multiple agents playing. As the current research on *Odessa* treats it as a single agent in a duet collaboration with a human, the system behaviour is thus designed to suggest a single agent.

- **Pitch algorithms.** Takes an array of 5 pitches as input; for each pitch, a pseudorandomly selected operator either lowers the pitch by one semitone, raises the pitch by one semitone, or leaves it unaltered. The three alternatives have a theoretically equal probability. The purpose of this transformation is to introduce slight variations, so that the module output is not identical to its input.

- **Velocity generator.** Generates a pseudorandom number within a MIDI velocity range of 50–127. Although MIDI velocity extends down to zero, in practice, when using *Odessa* with either synthesised output or a Disklavier, values below 50 are too quiet. The maximum possible MIDI velocity value is 127.

- **Velocity algorithms.** Incoming values are expanded into 5 pseudorandomly generated values, restricted to a range of 10 (integers) less than the incoming value, to a maximum value equal to the incoming value. Thus, if a value of 80 is received as input, output will be an array of 5 pseudorandom integers in the range of 70–80. If an incoming value is less than 60, the range of 50–60 is used. This 5-value array is transmitted as output. The purpose of this module is to introduce slight variations in loudness into the output stream. Output remains in a narrow dynamic range over short periods for small groups of notes, although drastic variations are possible from one note group to the next. This can give the impression of producing phrases with dynamic contours, while remaining responsive to sudden changes.

- **Time between pitches.** Finds the duration of silence between incoming notes, specifically between the end of one note and the beginning of another. This value was empirically determined to be more useful than the duration of notes, as it gives a sense of what could be referred to as sonic 'density'. Thus, whether staccato notes or long tones are received as input, the duration of silence in between notes suggests that more or less note activity is taking place. When the highest layer is activated, due to the module "Disregard input pitch", no new input arrives to overwrite the previous input, so the last used value remains in the register and continues to be transmitted until it is replaced by new input.

- **Sound process spawner.** Continuously outputs potentially overlapping 5-note segments of sound data. Can in principle produce synthesised output but, in the current implementation, produces MIDI output. If pitch value or velocity value are zero or not present, a rest (silent note) is transmitted. See section 3.2.1.1 above, especially Figure 1.

- **Throttle.** Inhibits pitch values (forces all notes to be rests) for an empirically determined duration of 500 milliseconds (.5 seconds) after each audible segment produced. In practice, this allows for enough overlap between segments to produce chords and complex rhythms (see Figure 1 above), but typically preserves the subjective sense of a single agent performing.

3.2.3.2 Modules in the *Adapt* layer

- **Pitch sensor.** Continuously polls the sonic input signal from human instrumentalist and analyses spectral information using native ChucK functions. Extracts the strongest frequency values from the spectrum and converts them to MIDI pitch values for transmission. Peak spectral information often picks out higher harmonics rather than the fundamental input frequency.

  This approach to input pitch analysis stands in contrast to a more computationally expensive procedure to more reliably pick out fundamental frequencies. A similar trade-off is described in Brooks (1999, pp. 43–44), where less computationally expensive sensor reading analyses, when used effectively, can lead to robust performance by a mobile robot. The practical aim of this module is to use the extracted pitch values to affect the pitch values in the system output, to facilitate collaborative interaction with a human co-performer. This aim is not compromised by picking out higher harmonics. In fact, this approach to pitch extraction actually gives the impression of an enhanced musical behaviour, by producing appropriate responses to richly harmonic input. In short, it facilitates the agent's sharing of a harmonic space with the human co-performer. This is accomplished with the Subsumption approach, that is, without recourse to any high-level formal knowledge of musical theory.

- **Loudness sensor.** Uses native ChucK function to extract peak RMS amplitude from the sonic input signal. In a pre-performance sound check, a maximum and minimum RMS amplitude of the input signal (for a given human performer) is mapped to a range of 7-bit values, where the maximum is 127 and the minimum is 50. Values 0–49 are treated as silence, that is, as if there is no audible input. In practice, this means that barely audible sounds (rather than strictly inaudible ones) may fall below the threshold, which, in practice, does not typically compromise system performance. These ranges were established in relation to *Odessa*'s output capacity, for which it was empirically determined that values below 50 are too quiet when using either synthesised output or a Disklavier.

3.2.3.3 Modules in the *Diverge* layer

- **Disregard input pitch.** Inhibits pitch data from human co-performer. Remains active for a period of 1–5 seconds, based on the expiry of a timer set by restricted pseudorandom number generation, reset on each activation so as to be irregular and unpredictable. (An additional effect of this module's activity was described above; see module "Time between pitches", from the lowest layer.)

- **Rest.** Suppresses loudness data from human co-performer, which in turn suppresses data from the velocity generator module, resulting in no audible system output (MIDI velocity values of output are set to zero). Remains active for a period of 1–5 seconds, based on the expiry of a timer set by restricted pseudorandom number generation, reset on each activation so as to be irregular and unpredictable.

- **Initiate ending.** In simple terms, this module tentatively stops the system from producing output. This represents a divergence from the human co-performer in so far as the system stops playing beyond the duration of an ordinary rest, similar to how a human co-performer may conclude a performance by dropping out. If all performers subsequently drop out, the first performer to do so may be considered to have initiated the ending (although human performers may also initiate endings with continued performing, rather than merely dropping out). This module only *initiates* an ending, because if the human co-performer continues rather than also dropping out, the system will resume operation. Specifically, when no input from the human co-performer is present, the module blocks system output (MIDI pitch values of output are set to zero) for a period of 5–10 seconds (based on the expiry of a timer set by restricted pseudorandom number generation, reset on each activation so as to be irregular and unpredictable). If human (sonic) input is detected during this period, normal operation resumes. If no input is detected during this period, the global system ceases operation.

This module, in principle, allows for a co-performer with knowledge of it how it works to force an ending by ceasing to play altogether; this would likely force an ending to a performance in a strictly human collaboration as well. With no input, the system could potentially continue to produce output for a maximum of approximately 15–20 seconds (assuming the maximum timer value is set in the *Play* layer, and some completion time is allotted for any already spawned segments). This module also has a single-use timer, activated only at the start of the system (i.e., the start of a performance), that blocks the module from activating for a fixed period of 30 seconds. This allows a human co-performer, before playing a note, to listen to the system play for a short time without having it come to an abrupt halt.

3.2.4 On the development of the initial prototype

The *Play* layer was the first to be implemented. Following the decomposition presented in Figure 2, it was determined that individual notes could be muted if either pitch or loudness information (velocity in MIDI) was omitted in the pathway to the *Sound process spawner* module (see Figure 3). However, even with some notes muted, the resulting output did not cohere as a steady stream of notes and chords. This led to the addition of the *Throttle* module, which could produce more muted notes by further inhibiting pitch information. The next step was to set the throttle to a sane value. If the throttle delay was too long, the individual fragments (the upper staves of Figure 1) were too spaced apart to produce a steady output stream. Conversely, if the throttle delay was too short, the stream was too dense and did not cohere as a steady stream of notes and chords, as with the output prior to the addition of the throttle.

Once a satisfactory balance was achieved with the throttle, the *Adapt* layer was implemented. The main requirement was to receive appropriately calibrated input. This was to be decomposed into pitch and loudness information that would suppress the pseudorandom pitch and velocity generation of the *Play* layer. I produced my own instrumental input into the system during this phase of development, using a rapid prototyping approach, followed by fine-tuning.

Finally, the *Diverge* layer was added to produce additional behaviours in the presence of input, including the ability to ignore the input (*Disregard input pitch*), to occasionally fall silent (*Rest*), and in some cases, to present opportunities for the performance to end (*Initiate ending*). With the addition of this layer, the system was not left to act merely one way in the presence of input, and another way in the absence of input. Further fine-tuning with my own instrumental input took place in this stage, until I determined that the system was robust and flexible enough to be used in the first round of studies with the participants.

The next chapter describes how *Odessa*'s construction and interaction model relate to the experimental design, and a detailed account of the experimental setup and methodology is given.

Chapter 4

Evaluation

In research on computer systems, broadly speaking, system design, experimental methodology, and evaluation are closely related. Section 2.6 described how this relation was theorised by key figures in the history of AI, in particular, by Alan Turing and Norbert Wiener in the mid-Twentieth Century; Rodney Brooks, whose Subsumption has been central to the present research; and, Joseph Weizenbaum, whose ELIZA has been frequently referenced in the literature on AI. Each of these thinkers have considered what a system design is intended to achieve, and how the conditions of the system's deployment are wholly relevant to the evaluation of the system. In other words, an evaluation considers what a design set out to achieve, but it assumes the implementation is deployed in the intended target environment. This background is important, as it reflects a critical understanding of evaluating interactive systems in a way that differs from other common modes of evaluation, especially of computer music systems.

The above-mentioned historical examples are relevant for understanding the methodological considerations and experimental evaluation of *Odessa*. For *Odessa*, what is being evaluated is the system's ability to facilitate a collaborative performance with expert human improvisors. To account for this evaluation, this chapter describes the methodological details of the empirical research on *Odessa*.

4.1 Methodology

4.1.1 Experiment Description

The experiment was designed to maximise ecological validity by matching a number of real world conditions. In this case, gathering the data 'in the wild' was precluded by the nature of what was being investigated, namely, the potential of a musical collaboration to be experienced by a human co-performing with a particular computer system. While Chamberlain et al. (2012) state that there is no clear definition of 'in the wild' research, the central objects of such research are typically integrated with everyday life contexts. The experiment was, however, designed to preserve many aspects of a relatively common mode of encounter among the international community of free improvisors, namely, when players who have neither met nor heard each other play engage in real time improvisational musical collaboration. It is also common within this community for a fellow performer or admirer of the music (i.e., someone with a sympathetic ear) to serve in the role of making an audio recording of a free improvisation, which was the role I served in the experiment. On the other hand, for the present study, I was known to the participants as the designer of the system, and they were aware that they were participating in an academic study. Both of these facts detracted from the ecological validity.

As Hammersley (1993, p. 433) states of qualitative research methodology, "there are no overwhelming advantages to being an insider or outsider. Each position has advantages and disadvantages, though these will take on slightly different weights depending on the particular circumstances and purposes of the research". In this case, I know each of the study participants on a personal basis, and have performed collaborative instrumental musical improvisation professionally with nearly all of them (unrelated to the system under consideration in this study). Knowing the potential participants personally and as improvisors made it possible to consider players from a wide range of backgrounds, instruments, and approaches to improvisation for inclusion in the study.

The selection process was guided by the aim of challenging the system with a heterogeneous set of interactions and garnering diverse perspectives on it. Having participants who are experienced and knowledgeable in discussing improvisation was also important; its success as a selection strategy partly depended on the participants' trust of the researcher as a conversation partner when speaking about a practice that is notoriously difficult to address verbally. More specifically, with knowledge of the difficult-to-articulate subtleties and complexities of contemporary musical improvisation, I was able to recognise provisional statements (which pose a risk of being misconstrued by those outside of the field), and to elicit clarifications and additional feedback that may have otherwise gone unstated.

A key disadvantage of my role was the sense that, given the (correct) perception that I was the system designer, the question remained as to how critical the participants could be while still feeling tactful and comfortable, in light of my potential discomfort during such critique. This raised the issue of the degree to which they might be holding back more critical responses. Two interrelated strategies were used to mitigate this disadvantage, incorporating modified "think-aloud" sessions and follow-up interviews (for a detailed account of traditional think-aloud methodology and a modified approach, see Koro-Ljungberg et al. in press).

4.1.1.1 Format and Procedure

The first strategy was to use unstructured verbal (modified think-aloud) protocols that took place immediately following the musical improvisations with the computer player, all of which preceded any discussion of the system by the researcher. This lack of discussion was significant to the framing of the improvisation, so as not to solicit any specific playing strategies that could implicitly guide the system performance and in turn influence the verbal feedback. The openness of the situation allowed for a wide variety of performance practices and reflections on personal experiences of the improvisations. It is also important to note that the name *Odessa* was not disclosed until the conclusion of the study (with the initial prototype).

Related studies of improvisation without computers have been conducted without a connection to a specific performance (e.g., MacDonald and Wilson 2006), or have used listening to recordings as a means for improvisors to reconstruct internal mental narratives of their performance (e.g., Sansom 1997). For the present study, it was more relevant to elicit immediate post-performance impressions of the participant experience, in order to focus on the ways in which specific performances were experienced. This latter form of commentary permitted considerations of the performance that likely would have been precluded by a linear analysis of musical playback. In particular, rather than moving across the temporal axis of the performance (i.e., section by section, from start to finish), the responses instead moved from more immediate thoughts to more reflective ones, and tended to oscillate between describing general aspects of the interaction and specific moments or sections. This more immediate consideration of specific performances thus offers one way to examine the experience of performing.

After three uninterrupted performance and verbal protocol sessions, a semi-structured interview was conducted. The interview questions were formulated to prompt long explanations and avoid implicitly suggesting a specific answer (see Stock 2004); this comprised the second strategy to encourage forthcoming critical responses. Thus, in place of asking, for example, "Did the system respond adequately to your playing?", the preferred formulation would be, "Did the system respond to your playing adequately, inadequately, or somewhere in between?". When apparently superficial or vague answers were encountered, follow-up questions helped to elicit more specific data (e.g., "You stated that the system responded to your playing 'pretty adequately'. How would you characterise what was *inadequate* about its responses?").

After completing all the individual sessions, the data was analysed as depicted in Figure 4. Verbal data describing *internal* mental or bodily states was analytically correlated across participants; verbal data about *externally* observable aspects of the improvised performances was correlated with the speaker's musical audio recordings of the improvisations. Additional interrelations were examined between these complementary data sets. The analysis is further described below, in section 4.2.3 (Detailed data analysis procedure).

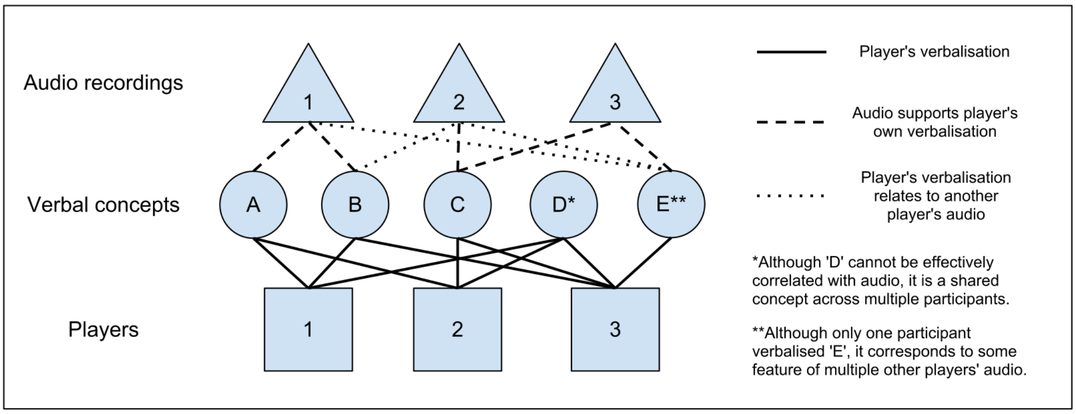


Figure 4: Data relationships.

4.1.1.2 Participants

The study consisted of eight case studies, each with a different performer and instrument. Those who participated are distinguished improvisors of international stature, who generously shared their time and expertise. The performers (five male, three female) have diverse backgrounds and span an age range of over three decades. To indicate the level of expertise, the variety of instruments, and the different approaches to improvisation, the participants are listed here (in alphabetical order by surname): Paul Cram, clarinet; Peter Evans, trumpet; Okkyung Lee, cello; Evan Parker, soprano saxophone; John Russell, guitar; Sara Schoenbeck, bassoon; Pat Thomas, piano; and Ute Wasserman, voice. At least four of them have had prior experience with interactive computer improvisors, though in two cases, not since the 1980s. In recent years, Parker has performed with a number of different systems, and Evans performed with an early experimental partially-automated Disklavier system I designed that later informed one design component of the initial *Odessa* prototype (musical stream decomposition).

4.1.1.3 Apparatus

The audio for the cello, guitar, and bassoon was captured using pickups that were largely impervious to audio feedback from the system output. This made for a clearer picture of the system's specific responses to player input. For the remaining players, a directional microphone was used, but despite careful setup, it did not achieve perfect separation between acoustic instrument and amplified computer output. Thus, at points when the system reached higher volumes, some of its output audio was introduced into the player's microphone as low-volume input. Because it was not possible to obtain higher quality directional microphones for the study, having the players use headphones was considered as an alternative solution. Ultimately, it was decided that using headphones would be too dissimilar to an ordinary playing situation, and would thereby compromise the overall experimental setup. It was thus determined that having the system respond to less pristine player input, somewhat contaminated by the occasional intrusions of audio feedback, was more preferable than having an atypical performance setup.

Consistency across studies was important to ensure a clear interpretation of the data, which would have been undermined by varying the sonic output mechanism. Thus, a self-imposed limitation of using amplified software synthesis was chosen, due to participant logistics and the practical difficulty of access to an electromechanically-controlled acoustic piano (e.g., a Disklavier) for all studies, although this would have been preferred. In comparison, a performance I gave using a Disklavier with the system resulted in what I found to be a more effective collaboration than previous interactions with the system's synthesised piano output (*Interactive Keyboard Symposium*, Goldsmiths, University of London, November 2012). The same Disklavier was used for a follow-up study with a second iteration of *Odessa*, described in Chapter 6.

Discrete pitches and an emulated piano timbre were used in the initial study to provide a familiar point of continuity and interrelation to the participants' previous experience. This was intended to help shift the verbal feedback to the topic of collaborative playing, rather than exploring the seemingly unbounded possibilities of computer-generated sound. Notably, however, from a technical perspective, the core of the system is easily adaptable and extensible to other input and output mechanisms. In particular, for input and output, it is currently capable of continuous as well as discrete pitches, and it is also possible to extend the system by taking timbre into account, without compromising the fundamental architecture (for a computer free improvisation system focused on timbre, see Hsu 2010). These options were deliberately excluded from the study to maintain its overall consistency and focus.

4.2 On the analysis

Qualitative analysis has the potential to offer a complex understanding of the subjective participant experience with *Odessa*, which is the phenomenon under investigation in this study. However, this complexity increases the difficulty of interpreting the data. Discretion is called for as to how exactly to interpret participant statements, which varies according to context. While one statement may be taken at face value, the next may be laden with implicit meanings that must be drawn out.

For example, one participant stated between two performances that "hopefully, I can be as creative as the computer", referring to the next improvisation in the series of three for the study. This statement cannot simply be taken to mean that the participant found the computer more creative than his or her own contribution. Using the clues of semantic context and tone of voice, I interpret this remark as being suggestive of the participant's ambivalence toward interacting with a machine, and judging it in human terms. In addition, it can be interpreted as a self-deprecating remark that relieves social tension but, more significantly, highlights the artificiality of the playing situation (an academic experiment). Thus, while a transcribed remark may initially appear, at face value, as a positive remark that would bolster the assessment of *Odessa*, its accurate interpretation relies upon the entire experience of the session, and is not necessarily self-evident in a single transcribed quote. I have taken great care to maintain a critical distance in the interpretive analysis, to remain sensitive to the diversity of meanings, positive or negative, reflected by the participant statements, as shown in the following two chapters.

When conducting the studies with some participants, in some instances, I encountered a degree of reluctance to reveal or even to think about negative or critical feedback. Many expressed a sensitivity to my feelings, knowing that I had designed the system, despite my repeated reassurances that positive and critical feedback were equally valuable to me. In my own strategic response to this, I found that I was able to elicit a large number and wide variety of critical responses through a combination of techniques, which included asking follow-up questions about comments made during the verbal protocols prior to the discussion; posing questions effectively; and, asking the right questions at the right time.

At times, participants' reflective, thoughtful responses indicated that positive and critical sentiments could be intertwined. Rather than a forced reductive simplification, a sensitivity to this complexity remains throughout my analysis, as demonstrated in the following two chapters. By preserving apparent contradictions in the data, the research value of the information is enhanced, because it can serve to form a more accurate understanding of the phenomena under investigation, instead of skewing the data by forcing it into preconceived conceptual categories.

Interestingly, out of eight participants, only one participant had an overwhelming majority of clearly positive assessments, and only one had an overwhelming majority of clearly negative assessments. The remainder had a mixture of positive and negative assessments. When similarities arose across a majority of participants about a positive or negative point, given the two outliers, it was usually possible to provide at least one contrasting statement. By providing a contrasting statement wherever possible, I am able to present a more balanced account of topics with wide agreement among participants. This further supports the impartiality of my account and indicates the successful minimisation of confirmation bias in the study.

4.2.1 Understanding the transcriptions

The transcriptions of the verbal protocols and interviews use the following conventions:

For readability, some filler expressions have been omitted, such as “um” or “uh”, but common filler words that may have relevant semantic connotations such as “you know”, “I mean”, or “kind of”, have been kept in.

False starts that are completed as they are initially indicated have been simplified, for example, “I (uh), I think that” would be simplified to “I think that”, but false starts that indicate a revision in formulation have been kept in. A hesitation and reformulation are indicated by a single dash connected to a word, for example “I feel that- I think that [...]”.

A significant pause, beyond the established rhythm of the participant’s contemplative speech is indicated by [pause] (in square brackets); pauses that exceed the established [pause] duration are indicated with [long pause]. Laughter is indicated with [laughs].

In some cases, a participant will add a comment that is inconsequential to his or her train of thought. For example, “I think that my response was — let me unplug this microphone — motivated by excitement”. Such instances are elided with ellipses in square brackets: “I think that my response was [...] motivated by excitement”. Ellipses in square brackets are also used to elide longer portions of a quote addressing a different topic, which have been omitted in order to condense several statements on a single topic. This is common when several topics are interspersed throughout a longer participant reflection; statements on a single topic are extracted and grouped together during the analysis.

If a comment is signalled as an aside, but is relevant to the sentence, it is included in between long dashes with a space on either side. For example, “I think that my response was — how can I say this? — motivated by excitement”.

The sense of the participant’s speech will attempt to be replicated as closely as possible using standard punctuation and italics for emphasis, for example: “I thought to myself, ‘oh my god, really?’, because I couldn’t believe *that* could be the case”. Colloquial expressions have been kept in; in some instances, these require special orthography, such as “because” shortened to“’cause”, and “going to” shortened to “gonna”.

Occasionally, in a long passage, a blank line is used to indicate the beginning of a new thought, for readability. Without “[...]”, “[pause]”, or “[long pause]”, it does not indicate an elision or rhythm of speech.

If a comment was made during the interview portion that was significantly shaped by the question, the question is included in ***bold italics***; if the answer is reasonably independent of the question, the question will not necessarily be included. Additional words may be inserted in square brackets to provide additional context as needed. For example, a question such as “Did you find the second improvisation interesting or uninteresting?” (in context, printed in bold italics) would be provided along with a statement such as “I found the second improvisation interesting”. However, a question such as “What did you think of the second improvisation?” may not necessarily be provided with a statement such as “I found it more interesting than the previous one”. Instead, the answer may be given without the question as “I found [the second improvisation] more interesting than the previous one”.

Finally, it should be noted that the full speaking context includes physical gestures, speech inflections, etc., as well as the much broader semantic context established by the entire meeting. In rare instances, these details, not captured by the verbal transcript, may allow for a specific meaning to be inferred that is not apparent in a given extract. I will supply additional words in square brackets only when I have a high degree of confidence about the understanding, based on the full context. An imagined example might replace:

“The way I was- right there, I did that flourish — you know? — and it did that loud thing”.

with:

“The way I was [playing] right there, I did that flourish [of notes] — you know? — and [the system] did that loud thing”.

In the above example, the most significant clarification is the referent of "it". The correct referent of "it" may have been inferred from a physical gesture or a previous comment, for example, or may have been explicitly established by a follow-up question; I have not augmented any legitimately ambiguous comments with purely speculative inferences.

In my own references to *Odessa*, as long as a specific descriptive term is not misleading in a given context, I will use the proper name *Odessa* and the terms "system", "program", "computer", and related terms, interchangeably (e.g., "improvising with [*Odessa*/the system/the program/the computer]"). As stated, the name "Odessa" was not disclosed to the eight participants until after the conclusion of the study (with the initial prototype). At some points in the transcriptions, however, I have inserted "[*Odessa*]" (in square brackets) into their comments for clarity. As discussed further below, participants typically varied between referring to the system as "he", "she", and "it", as well as the "system", the "program", the "computer", and related terms. One participant assigned *Odessa* a common female name that, while unlikely, could potentially indicate the participant's identity; in the participant's comments, this proper name has thus been replaced with "[she]" or "[her]" (in square brackets).

4.2.2 Anonymising the data

Although the identities of all eight participants have been disclosed in order to reveal the level of expertise and diversity of backgrounds, their names have not been directly linked to their comments. Instead, names have been replaced with anonymised unique identifiers. In this way, a set of participant comments can still be attributed to an individual participant, while preserving the anonymity of the contributor. All references to what instrument was played and to other named individuals, musicians or otherwise, have been redacted.

There are a number of reasons for this anonymisation, the first of which is to underscore the fact that the study was designed to scrutinise *Odessa*, not any of the participants. The participants have each generously responded to the demands of the study with their individual perspectives, which have been equally valued here. Also, comments linked with particular names could risk unforeseeable personal or professional harm to a given participant.

Another reason for the anonymisation of comment attributions is that, during the experimental case studies, I provided a space for musicians to speak casually and colloquially, as I, too, generally speak in everyday conversation; they were not called upon to make academic formulations, nor were they given time to reformulate extemporaneously spoken dialogue into a written form. For the purposes of academic presentation, however, I am indeed called upon to make academic formulations, especially academic reformulations of colloquial speech taken from these studies. This task of clarification (and correlation among participant views) should not be viewed as an improvement on (and thereby a diminishing of) anyone's comments as they were originally spoken.

4.2.2.1 Anonymisation procedure

A list of eight surnames, arbitrarily selected from a list of English surnames, was initially generated according to the criteria that the names should not strongly suggest any of the participant names, nor be obviously suggestive of specific musicians unrelated to the study. Surnames that commonly appear in citations in this thesis were also avoided, as were names ending in the letter 's' (to avoid awkward possessive apostrophes). These eight surnames were then shuffled into a random order, as were the eight participant names. The resulting two randomly-ordered lists were then paired in sequence, resulting in a substitute surname by which the participant's quotes are identified. To reiterate, the substitute names do not stand in any meaningful or decipherable relationship to the actual names. The pseudonyms are: Anderson, Campbell, Hamilton, Johnston, Morgan, Quinn, Stewart, and Walker.

4.2.3 Detailed data analysis procedure

For the research on *Odessa*, the process of analysis proceeded through the following steps, as depicted in Figure 5. The notion of "top-down" is used to indicate that there are particular topics that I am imposing on or eliciting from the data; "bottom-up" is used to indicate that I am using a neutral analysis to discover topics that appear in the data. A more detailed description of the steps in Figure 5 is as follows:

(1) Concerns the construction of the experiment, which was designed to elicit commentary on topics relevant to evaluating the collaborative experience of improvising with *Odessa*. This is, to an extent, "top-down", for example, when I ask an interview question pertaining to the system's responsiveness, whether or not the participant has chosen to focus on that topic. At the same time, however, the experiment was constructed to allow participants to speak freely, and potentially diverge from the topics that I introduced.

Thus, in (2), certain data was expected to match the topics I introduced, but I also examined the data for recurring themes across participants that were not explicitly solicited.

In (3), the themes that were positively identified (whether or not they were expected) were then organised according to overarching thematic groups: behaviour-related issues,

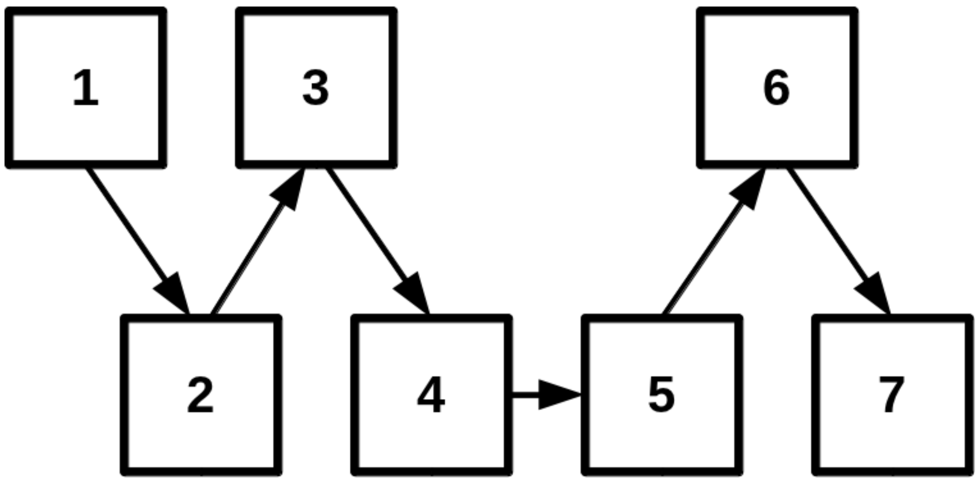


Figure 5: 1. Experiment design to capture data for research question (top-down);
2. Analysis of emergent themes (bottom-up);
3. Re-organisation of data according to emergent themes (top-down);
4. Excluding of irrelevant data (narrowing);
5. Analysis for salient experimental results (bottom-up);
6. Re-organisation of data for presentation (top-down);
7. Presentation.

music-related issues, and issues pertaining to the construction of the study itself. All of the data was then labelled according to themes within each of these groups.

In (4), I excluded off-topic comments that did not pertain to the participant experience, the system, or the study, as well as excluding comments that were unintelligible, usually due to participants composing their thoughts.

In (5), the remaining comments were studied along with the audio recordings of the musical performances, to better understand verbal references to musical experiences, and to verify comments about specific musical passages. In all cases, the participants' retrospective descriptions of improvisations proved accurate, so no verbal data was excluded at this stage.

In (6), upon further analysis of the organised materials, I reconceptualised the themes into five main (occasionally overlapping) topics, in order to help focus on the salient experimental results. At times, several of these themes may be represented in a single participant statement. Other times, a single theme may be sufficient to characterise a participant statement. The five themes are as follows:

Co-performer: anything directly related to the participant's sense that the agent is co-performing or collaborating with them, whether self-identified as originating in human psychology, or externally observed in agent behaviour;

Identity: anything concerning the participant's sense that the agent exhibits a unified set of behaviours, has a personality, or performs actions with intent, whether self-identified as originating in human psychology, or externally observed in agent behaviour;

Positive human comparison: anything concerning the participant's sense that the agent is similar to other human performers with whom the participant is familiar (these may be criticisms or accolades);

Negative human comparison: anything concerning the participant's impression that some attribute, skill, behaviour, etc. is beyond the capacity of any non-human; may be justifiable or due to an unfounded bias; and,

Music: anything concerning musical features of the agent's performance, including traditional musicological categories such as harmonic assessments, and also social-behavioural categories that relate to contemporary musicological theories of improvisation, such as how the agent acts in a given musical circumstance.

Finally, (7) is the presentation itself, which draws upon the focused themes from (6), but ultimately presents the data in a manner that best allows the individual participant experiences to come through, while at the same time, supporting my argument in a sustained manner. Thus, these themes will not necessarily be readily apparent in the exposition, but they were important to the analysis.

With respect to the point in Ball and Ormerod (2000) (cited in Chapter 2) that one could find a middle ground between strict ethnography and strict experimental science, I have already shown the extent to which I have constructed the experiment to effectively investigate certain phenomena. Regarding the ethnographic ideals, I have attempted to preserve the voices of the participants, allowing them to speak for themselves, to an extent. However, at the same time, any qualitative approach such as ethnography must also present the ethnographer's own evaluations, judgements, extrapolations, and arguments on the basis of the participant data (Holliday 2007). Thus, I will typically interject my analysis throughout the interspersed participant comments. There are some cases where it is valuable to allow comments to extend beyond the typical length, for the conveyance of a broader semantic understanding with respect to individual voices. When such longer quotes are used, I will always follow with an exegetical exposition.

The next chapter presents the analysed results and the corresponding evaluation of the system.

Chapter 5

Results

The central issue investigated by this empirical study is the extent to which human participants experienced the artificial agent, *Odessa*, as a collaborative co-performer in free improvisation. In the course of the study, this research topic has suggested several constitutive questions: Was *Odessa* perceived as a uniform agent?; and, Was it perceived to be engaged in a shared musical practice with the participant? There is also the issue of where these perceptions originate: What are the roles of the human participant's imagination and biases?; and, To what extent does *Odessa* facilitate or negate initial human presuppositions about it?

While it will not be possible to disentangle all of these questions in every case, the experiment has proved successful in at least one significant respect: it has elicited critical self-reflection by the participants, which has touched upon all of these issues. Occasionally, these issues may be considered in isolation, but, more frequently, they are discussed as intertwined or overlapping parts of a complex experience. This chapter will present the evidence that most participants did in fact experience *Odessa* as a collaborative co-performer, although, as shown below, the evidence does not offer a simple and direct answer. Guided by the above questions, this chapter will present different aspects of the complex experiential phenomenon of collaboratively improvising with the system. This evidence supports a key idea presented in previous chapters, namely, that the system performance cannot be located in the system alone, nor in the human perception alone, but rather, in a tightly coupled interaction between the two that involves both human psychology and the system's behavioural dynamics.

5.1 Introduction

On the basis of my analysis, I will determine to what extent the predictions of the hypotheses have succeeded. This will include a presentation and discussion of the strengths and successes of the system design, as well as its limitations or shortcomings. Where relevant, I will also address the effectiveness or ineffectiveness of the experimental setup. The presentation aims to provide the clearest account of the salient evidence for my arguments.

One complexity that became clear in the analysis is that participants may be *adapting their own behaviour to the system's, in order to produce a complementary result*, while, at the same time, *maintaining a sense that the system is responsive to their activity.* This responsiveness allows for participants to recognise that their co-performance with the system is fundamentally different from the activity of playing music to complement or accompany static output. The duality of the participants adapting their own behaviour to *Odessa*, yet perceiving its adaptation to them, suggests an analogue to the real-world situation of an expert performing with a competent but less skilful improvisor, in a way that indicates that the expert values the performance and respects the other performer. One aspect of the results, discussed below, suggest that *Odessa* can be understood as a competent but less skilful improvisor.

Many salient points discovered during the analysis are more clearly identifiable when they are permitted to emerge from a series of interconnected statements, even if such statements are not necessarily articulated by the participants as a set of discrete points. I will, however, excerpt participant commentary into intelligible units, which can range from a single sentence to several paragraphs. My on-going analysis is usually interspersed with the participant comments. At times, it is beneficial to present long passages of participant commentary; these are always followed by my commentary to help better understand the similarities and differences among the participant views.

5.1.1 Initial analysis I

I will begin by introducing Anderson's account. This account was chosen as an introduction due to the high concentration of central themes in a continuous, relatively short passage. This passage will thus serve as an introduction to a group of assessments that can be found, in a closely related form, throughout all of the participant comments, with one exception. The exception is for Johnston's account, which presents a significantly different experience with *Odessa* that is covered in section 5.2.2. With the exception of Johnston, the common points of assessment are drawn from Anderson's account and used to identify similarities and differences among the participants.

Using the analytical approach described in section 4.2.3, eight principal points have been identified in the transcripts, which are introduced here gradually, and then, further below, listed together. Initially, it will be shown how these points relate to Anderson's commentary, and then how they relate to the commentary of the other participants. To provide a specific way of reading the initial passages, some of the points are introduced in advance. Further below, the derivation of each of the points is shown.

The first two relevant points are:

(1) If the co-performer is a machine, it will fail to be a valid collaborative partner; and,

(2) Some human players ignore you; it is unpleasant to play with them.

A statement by Anderson, that describes a metaphorical understanding of the musical interaction experience with *Odessa*, offers an example of (1) in terms of a bias against the capabilities of a machine, and also contains a reference to (2) in terms of poorly-regarded human players:

> **Anderson:** I kept looking for something [in the music] that I can kind of go inside, 'cause- obviously it's [a musical mutual interrelationship is] not gonna happen, because it's not really a live person, even though there're kind of plenty of people who play like this [who do not permit a co-performer to interact with their musical material]. And after awhile, I kind of stopped looking for that but- because usually it's kind of fun to have that moment, to go under or go inside.

I followed-up on Anderson's expression of going "inside" or "under" someone's musical material:

> ***You mentioned the idea of — [after] the third [improvisation] — [that] you were missing a chance to be able to "go inside" or "go under". I wondered if you could just say a little bit more about that?***
>
> **Anderson:** Lots of times when you're improvising, you kind of, there would be a locking point, you know? It could be a melodic thing, it could be sound, or it can be a rhythmic thing or something, and you are not so sure where you're going with it. But you're kind of holding on to this idea, together —

> unless you're playing solo — and then you're waiting, you're trying to- you're not trying to push it, but you're kind of working on that little material until it goes into- it blooms into something musical. [...] and obviously you can do it with a live partner, who can kind of go with it, with you.

The above statement ends with another comparison to a human ("live") partner, although unlike the first comment above, this one seems to be based on the improvisations with *Odessa*, rather than based on biased expectations regarding any given computer's capabilities. The other participants have similarly referred to (1) in terms of the limitations they assumed of the computer. Their comments, while informed by their improvisations with *Odessa*, also, to an extent, indicate a bias held before the experiment began:

> **Hamilton:** It takes a lot of courage, and these are all — I mean it seems silly to repeat all this — but it's obvious that this program doesn't have any of these qualities, by itself, and maybe they reflect the person that made the program in the first place.

Elsewhere, Hamilton and others give a further indication of this bias (participant responses, given separately, are listed here together):

> **Hamilton:** When you're playing with it, you're wrapped up in anthropomorphising the program and also trying to forget about that, and it's easy to get caught up in sort of mind games, even if you try to play naturally or just do what you would normally do; even that is kind of a mind game, because *you realise that that's not really what this is about. It's not gonna react naturally, so why should you?* [emphasis added]
>
> **Campbell:** I don't know how you would make that [performance] really feel like you're improvising with another person, when you're [in fact] improvising with a computer.
>
> **Stewart:** Inevitably, I can't pretend that this is [like me] playing with another human being; this is playing with the thing that you designed.
>
> **Morgan:** It's certainly different than playing with a human, don't you think?

Another related but significantly different perspective is offered by Quinn:

> **Quinn:** I wasn't kind of thinking about my playing partner as being a machine really, at all. Or at least, I mean, I *knew* that it was a machine, but I was sort of playing as if it wasn't a machine, and I found it was quite a sort of human characteristic.

In Quinn's account, treating *Odessa* like a human in the musical performance led to experiencing *Odessa* as less machine-like, despite the knowledge that it was indeed a machine. A similar sentiment is captured in a longer passage from Anderson below.

To help understand Anderson's comments below, two additional points are relevant:

(3) Being completely engaged when improvising results in a favourably-experienced improvisation; and,

(4) There is an observable distinction between the experience of a non-collaborative and a collaborative performance.

Using the above points as a guide, it is helpful to read the following lengthy passage in its entirety, in order to get a broad sense of what is being conveyed. Following this long passage, I will analyse each segment of it in order to identify within it points (2)–(4) described above, as well as additional points (5)–(8).

Anderson elaborates on the idea about (2) in terms of poorly-regarded players, hinted at above, but also offers a favourable account of the experience improvising with *Odessa* that suggests (3) in terms of the mental space of a valid collaboration, and (4), a positive sense of parallel playing:

> **Anderson:** [In some cases,] you can just [say] "oh, that person 'so and so' doesn't listen to other players"; [...] I think there is definitely that case. And some people play in a way that you feel like whatever you're doing doesn't matter; it's not going to change anything of what they're doing and, in a way, it makes you feel like, 'well, what was the point of me playing with this person?'.
>
> But then, also, there are people, like, who just kind of develop ideas in totally different ways from you, but somehow it's- it cannot — I mean, this is not something you can understand logically when you're doing it, and maybe it's a different way of wiring in your brain — but somehow, like, as long as you, kind of, you can keep up with the other person, not trying to follow, but kind of keep up with it, somehow, then it becomes interesting. It doesn't become like two people playing almost in a separate room — you know what

> I mean? — somehow, it's a little bit more coherent music coming out, even though two of them are developing things in such different ways. [...]
>
> ***And did you ever feel that [coherence] with this [program] in any of the three improvisations? Anything in that direction, or was it just completely missing that feeling of coherence?***
>
> **Anderson:** I have to say, all the time I was feeling it. I mean, I didn't feel like- funny thing is, like [after the first improvisation], I was saying "no matter what I do, it's not gonna make any difference", but the thing is, like, I somehow knew that it *was* making a difference. I think [in] the second [improvisation], maybe I was kind of looking for it a little more, like trying to find something to connect to, or like latch myself onto its logic, or whatever. I mean- but then, that's when it happened that I kind of decided to kind of just go parallel, not trying to be on it or in it, and it felt actually quite good.

So despite a self-admitted bias against the machine's capabilities, a positive collaborative performance was experienced. This experience with *Odessa* in at least one way exceeded the expectations Anderson has of some human players, namely, those who appear to ignore their co-performers.

There are a number of elements in this long passage above that provide clear points of comparison to other participant views, both in agreement and disagreement. As these points are contained in the above passage, which has now been presented in context, it is now possible to consider them separately. I will refer to these below as three distinct paragraphs (the question, in bold, appears in between the second and third paragraphs).

5.1.1.1 Central cross-sectional relationships in initial analysis

For clarity, I will isolate and paraphrase each point in the above passage. The paraphrased points are presented in immediate succession, in order to allow for a quick comparison to the original passage above. While this set of points is concentrated in the comments of a single participant, Anderson, the points are tied to the other participants below, as well as to additional comments made by Anderson. Point (1), introduced above, is not relevant to the discussion below; however, points (2) – (4) from above are repeated, to show their derivation, and four additional points, (5) – (8), are introduced.

The first paragraph could be paraphrased as:

some human improvisation partners continue their activity as if entirely unaffected by your activity, which diminishes or renders displeasing the joint participation in a would-be collaborative activity.

More simply, this may be condensed to:

(2) *some human players ignore you; it is unpleasant to play with them.*

The second paragraph could be subdivided into two points, respectively paraphrased as:

there are some difficult-to-articulate aspects of the first-person improvisation experience that seem to involve a particular mental space or psychological state; while these aspects apparently resist an ordinary language description, they are nevertheless palpable and make for an enjoyable (or interesting or engaging) activity;

and:

there is a noticeable difference between juxtaposing the unrelated performances of individual players and coherently integrating parallel, independently developed musical materials in real time.

More simply, these can be respectively condensed to:

(3) *being completely engaged when improvising results in a favourably-experienced improvisation*; and

(4) *there is an observable distinction between the experience of a non-collaborative and a collaborative performance.*

The third paragraph contains three points, subdivided and paraphrased as follows:

(5) *the output of* Odessa *is affected by what I am doing*;

(6) *I can enjoy the experience of improvising with* Odessa; and,

(7) Odessa *seems to have its own approach (or logic or language) that I can observe and seek to relate to.*

A corollary to point (7) is drawn from the first two excerpts presented in this chapter (on the topic of going "under" or "inside"):

(8) Odessa *does not seem able to observe and relate to my approach (or logic or language).*

These points — or their negative counterpoints (e.g., in relation to (5), the output of *Odessa* is *not* affected by what I am doing) — recur among the other participants, as shown below.

5.1.2 Initial analysis II

The role of the above passage as a point of comparison for the others has only been assumed in the presentation. During the analysis, once the key points had emerged from a consideration of all the participant data, it became clear that a number of them were concentrated in the three paragraphs of a single participant, which facilitated this mode of presentation.

I will now establish the extent to which these points align with the views of the other participants. I will do so by citing passages that directly relate to these points, even if in disagreement with them. The passages will also be presented in a way that highlights specific language that recurs among participants. For convenience to the reader, I will use not only the numerical codes, but also repeat the simplified points in their entirety.

Hamilton states that:

> (**Hamilton:**) I don't feel so uncomfortable using these words, but- it takes, as a musician — to make certain types of decisions with another musician — [it] takes a certain amount of courage or just not really- a certain level of just not really caring, and just diving in. And what's fun about the sort of vibe of [*Odessa*] is that it definitely has that quality of just jumping in, which usually people have when they first start — if they're lucky, they get there just really excited to play. Or, there's a sort of different vibe you get from older, more experienced players where they don't really wait around for anything to happen — they just start. So, there's that.
>
> But then, also just the actual decisions it makes, which are really oblique sometimes and hard to figure out, and don't necessarily relate in a really obvious way to what you're doing, is- then that's some of the time [how *Odessa* plays]. Other times, actually, it related really obviously to what I was doing, so, all those things together made it a kind of complex personality which I would say makes it seem like a really experienced player.
>
> But then, there's also this — like I said before — there's almost a childlike quality to what the actual material is. So all these things together I think is what I mean as far as [my earlier comment that *Odessa* was either] novice or genius, or maybe some kind of combination.

In the first paragraph, we can see the suggestion that (7) Odessa *seems to have its own approach (or logic or language) that I can observe and seek to relate to.* In the second paragraph, there is the suggestion that (5) *the output of* Odessa *is affected by what I am doing.* The third paragraph has been presented here in context, but is revisited further below.

In another excerpt from Hamilton, below, the term "discouraging" points to (8) Odessa *does not seem able to observe and relate to my approach (or logic or language).* But the excerpt concludes with a more favourable assessment that suggests (3) *being completely engaged when improvising results in a favourably-experienced improvisation*:

> ***Did you find that it engaged you in what you were doing, or off and on, or not at all?***
>
> **Hamilton:** 'Engaged me' meaning 'encouraged me to do things'?
>
> ***Yes, sure.*** *[I did not necessarily find that Hamilton's reformulation of the question, above, was equivalent to the original question, but it was clear that Hamilton had a relevant point to discuss.]*
>
> **Hamilton:** Yeah- sometimes I find it discouraging, actually, but that's fine too because then it pushes you into some other area. But I found it engaging. It forced me to try things that I wouldn't necessarily normally try, and that's, for me, that's the whole point.

The overall impression given throughout these and other comments by Hamilton (presented further below) is that a collaborative experience with *Odessa* is perceived at some (perhaps intuitive) level, even without a reasonable explanation, on the basis that (4) *there is an observable distinction between the experience of a non-collaborative and a collaborative performance.* The question of what engenders that perception, and to what extent its source resides with the human, the computer, or both, is explored further below. Concerning point (6), *I can enjoy the experience of improvising with* Odessa, Hamilton states that "I found it to be really entertaining, and then kind of fun".

Some of the language used by Hamilton (above) coincides with language used in Stewart's passages below. (The below passages also tie in with other core themes, first identified in Anderson.) Stewart states that:

> (**Stewart:**) I would say [*Odessa*] was like an intermediate player, but with a good attitude ready to take a few risks. Good- not especially adept at [pause] intuiting the significance of repetition- or, there were a few times when, having identified a pitch sequence from [...] [*Odessa*], I used that same

pitch sequence later on to see if it would produce any kind of memory-like response, memory based. Do you know what I mean?

And you found that it didn't?

Stewart: I found that it didn't. It was much concerned with the future, which is kind of good. There's a kind of bright optimism about it which you could- is almost childlike. But, it's got some skills which are not childlike. So it's- yeah, to describe it as intermediate is just the start, you know. You could say in some ways it's rather sophisticated and, in other ways, a little bit dim, or lacking perception.

This primarily points to (8) Odessa *does not seem able to observe and relate to my approach (or logic or language)*, but also has an interesting similarity to Hamilton in how the system's apparently unlikely combination of strengths and weaknesses is perceived, and the notion that it is, in some respects, "childlike". Stewart offers additional comments regarding point (8), but also introduces other points summarised below:

Stewart: When I respond to it, it should respond to the fact that I responded to it, rather than just move on as though it's playing me. I don't mind up to a certain point, playing with musicians who behave like that. But, after a certain while, I start to get irritated you know — are we playing together or not? — and this experience [with *Odessa*] gives the feeling that there's [pause] not an adequate reward for the attention that I'm paying to it. It just grinds on with its own agenda. Which is, of course, like playing with — [...] I won't put the name in, but there are [human] musicians who play like that as well.

The below comment by Morgan is, in some sense, an inversion of the first sentence of the above comment by Stewart, but the implications are ultimately the same. Morgan states that:

(**Morgan:**) What happens is, it provokes you to sound like- you do something and it sounds like what you're doing, and then you think you should do that with them [*Odessa*], so you're responding to them responding to you — you know? — which not all the time is that interesting.

(The grammatical use of the singular "them" in the above passage is discussed further below.) These comments further emphasise the point that (8) Odessa *does not seem* *able to observe and relate to my approach (or logic or language)*, and also bring up the fact that (2) *some human players ignore you; it is unpleasant to play with them.* They also point to (7) Odessa *seems to have its own approach (or logic or language) that I can observe and seek to relate to*, although for Stewart and Morgan, the attempt to observe and relate to *Odessa* does not seem sustainable. For them, the experience was thus not sufficiently engaging, which resulted in an unfavourably-experienced improvisation. In principle, this could have been counteracted on the basis of (3) *being completely engaged when improvising results in a favourably-experienced improvisation*, but the above comments suggest it was not.

However, other comments by Stewart provide a more balanced view of the experience than is suggested by the above comments alone:

> ***Overall, would you characterise the nature of the duets as having a mutual collaborative result, or did it feel more like you were a soloist being accompanied, or vice versa?***
>
> **Stewart:** Oh, no, that's not the way I would play anyway. It felt collaborative, yeah, but to what extent this young lad can collaborate, that's what we're talking about. But there are plenty of human players that don't have the [pause] chops to make an adequate response, if I can put it like that.
>
> ***So you felt it adequately responded?***
>
> **Stewart:** Yeah, I enjoyed it. [...] I wasn't thinking, 'oh my god, what am I wasting my time here for?' [...] Not, 'oh, this thing is useless', or 'this is an uninteresting experience', not at all. This is interesting, and engaging, and demanding, even.

These comments suggest that (5) *the output of* Odessa *is affected by what I am doing*; (4) *there is an observable distinction between the experience of a non-collaborative and a collaborative performance*; and (6) *I can enjoy the experience of improvising with* Odessa.

The remainder of this section will present comments by Campbell and Walker. It is worth mentioning that this section will continue to use the numbered points to highlight the cross-sectional relationships across participant data. There is a transition, however, in the following section, to a more fluid style of presentation.

Campbell expresses a sense of both (7) Odessa *seems to have its own approach (or logic or language) that I can observe and seek to relate to*, and (8) Odessa *does not seem able to observe and relate to my approach (or logic or language)*, in the following passages:

> ***Did you feel that, out of the three performances, did you feel that there was a mutual collaborative musical result, or did it feel more like a soloist / accompaniment scenario, or how would you describe that?***
>
> **Campbell:** I would say, I mean it was definitely a duo, but I felt like I was [pause] bending towards the computer's language more than the computer bending towards my language, which is fine. You have to do that sometimes in performance [with human musicians] too.

These additional comments by Campbell give a sense of (2) *some human players ignore you; it is unpleasant to play with them*, and also suggest (6) *I can enjoy the experience of improvising with* Odessa:

> ***In terms of your experience throughout the three performance, did you feel like that it engaged you off and on, not at all, a few times-***
>
> **Campbell:** Yeah, it engaged me off and on, for sure. Yeah, absolutely. And it was interesting, like, anytime you improvise with somebody- it was interesting. I felt like I was able to come up with a common language by the third improv[isation], so it felt more successful to me that time.
>
> And it was- what's interesting, too, is that it was less boring than improv[isation]s that I've done with people, where you're just like, 'c'mon' [...], you know, like- it was more interesting than even like these two people that I played with recently, where I was just like [feigns exasperation], it was-
>
> [With *Odessa*], I would say it was not painful. It was- it did give me room to have ideas and so that's kind of amazing to think about, that this computer was more interesting to play with than this [person] I played with last month.

A related point to idea expressed above, that "I felt like I was able to come up with a common language by the third improv[isation]", is given by Walker:

> **Walker:** Well, I guess I got a bit used to it, because [...] [by the time I reached] the third one [improvisation], I found out how I could connect with it in a way that I find interesting. Like, I think, with an improvisor, you always have the meeting- a point where you're meeting, or [you] communicate [with each other] with your music.

Here, again, there is a clear sense of (7) Odessa *seems to have its own approach (or logic or language) that I can observe and seek to relate to* and there is also the suggestion that (8) Odessa *does not seem able to observe and relate to my approach (or logic or language).* Continuing with these two points, it is interesting to note how Walker's comments relate to one of the original comments by Anderson that was presented above; reproduced here, Anderson states:

> (**Anderson:**) Lots of times when you're improvising, you kind of, there would be a locking point, you know? It could be a melodic thing, it could be sound, or it can be a rhythmic thing or something, and you are not so sure where you're going with it. But you're kind of holding on to this idea, together — unless you're playing solo — and then you're waiting, you're trying to- you're not trying to push it, but you're kind of working on that little material until it goes into- it blooms into something musical.

Walker expresses more on this idea of development:

> **Walker:** There's a lot happening with- what's great is, I think that I feel like I can really improvise with it [*Odessa*], and find something, and get feedback from the program, and- I develop ideas while I improvise because I get that feedback. But, yeah, something about the timing, like- I find it's also difficult [with *Odessa*] to find the end of the improvisation, which is happening when you, yeah, improvise live — that you have that magic moment, or you don't have it. I mean, it's not there in every improvisation, that you have that moment: that you find an end together; or shape phrases together; or find a silence; or, yeah, whatever. [...] That timing you develop, or, I would develop, with a live improvisor is different- it's different with that program [*Odessa*].

Campbell also makes a similar comment, related to the co-development of musical material:

> **Campbell:** I would have maybe wanted a little more [pause] kind of like *unfolding* of the music. I felt like I could have used more.

So there is a recurring notion of the development or co-development of musical material, that has also been described as "blooming" or "unfolding". This point could potentially be codified as (9), although it did not seem to stand out in the data as strongly as the other points. It is mentioned here, however, as it seems to be important to understanding the system behaviour. The system behaviour is further considered in section 5.2.

5.1.3 Summary

To summarise the above section, it seems clear that many of the participants share a view of how they experience improvisation, both in collaborative situations, and in non-collaborative situations. Many of them distinguish between a co-performer that is completely ignoring them, and one who is actively involved in the collaboration. There is also a distinction drawn between a generally behavioural collaborative engagement, and a more specifically musical co-development of material. While several have noticed that *Odessa* does not completely ignore them, most felt that they were also conforming to its musical or behavioural tendencies more than it was conforming to theirs.

The following table gives a rough depiction of how the themes are represented among the participants in the above section, though it does not reflect the quantity of comments. The themes are given in the left-hand column, and the participant initials are given across the top. An 'x' is used to indicate that the theme is supported by the participant's comments, while a '-' indicates that the participant was referenced as support, even if the comment provided a contrast to the theme.

Names: Anderson, Campbell, Hamilton, Johnston, Morgan, Quinn, Stewart, Walker

Themes:

1 If the co-performer is a machine, it will fail to be a valid collaborative partner.
2 Some human players ignore you; it is unpleasant to play with them.
3 Being completely engaged when improvising results in a favourably experienced improvisation.
4 There is an observable distinction between the experience of a non-collaborative and a collaborative performance.
5 The output of *Odessa* is affected by what I am doing.
6 I can enjoy the experience of improvising with *Odessa*.
7 *Odessa* seems to have its own approach (or logic or language) that I can observe and seek to relate to.
8 *Odessa* does not seem able to observe and relate to my approach (or logic or language).
9 *Odessa* does not develop or co-develop the musical material.

	A	**C**	**H**	**J**	**M**	**Q**	**S**	**W**
1	x	x	x		x	-	x	x
2	x	x			x			
3	x		x		-	x		
4	x		x				x	
5			x				x	
6		x	x				x	
7		x	x		x			x
8		x	x		x		x	x
9	x	x						x

Summary table of participants and themes.

In the next section, I will build on the previous section, but attempt to provide a more fully elaborated account of the data. I will explore the concepts of *Odessa*'s perceived predictability and perceived responsiveness, while also considering the role of the participant's perspective in performing with the system and interpreting its actions. A related issue is the perceived contrast between performing with a non-responsive or static system, and performing with *Odessa*.

Following the more overtly systematic presentation in the above section, for the remainder of the chapter, I do not make repeated reference to a set of numbered points. I do, however, continue to present a clear analysis and development of themes. In the more fluid presentation style that follows, I present a series of themes that transition into others, without necessarily returning to the initial themes. While the analyses above have introduced both the analytical procedure and the data, now that the analytical procedure has been established, I believe the following will provide a more thorough understanding of the data.

5.2 Assessment of system behaviour

This section assesses aspects of *Odessa*'s behaviour on the basis of an analysis of participant commentary that, for the most part, has not been covered above. It considers the system's responsiveness (or lack thereof), the experience of entering a "zone" during an improvised performance with the system, and several concepts related to the experience of collaborative performance with the system. It concludes with a summation of this assessment, with a focus on what has been found to be relevant to the next iteration of the design, discussed in the following chapter.

5.2.1 Responsiveness

In the previous section, Campbell stated of *Odessa* that "it did give me room to have ideas". This indicates an indirect role that *Odessa* served. A slightly more direct role of *Odessa* was experienced by Stewart, who stated, "that's not the way I would play anyway", which (in context) suggested the meaning that his or her playing was influenced by *Odessa*. An even more direct role was accorded to *Odessa* by Hamilton, who stated that "it forced me to try things that I wouldn't necessarily normally try". This more active sense of *Odessa*'s influence was also perceived by Quinn, who spoke at length on the topic:

> ***Did you find that [*****Odessa*****] was listening to you and responding to you, or did you find that it was just doing its thing, and you were doing your thing, on sort of separate, parallel tracks?***
>
> **Quinn:** I thought she was great actually. I, you know- when you're improvising you- [...] you don't really think sort of in a partic- necessarily in a particular analytical way. I wasn't kind of thinking about my playing partner as being a machine really, at all. Or at least, I mean, I *knew* that it was a machine, but I was sort of playing as if it wasn't a machine, and I found it- it was quite a sort of human characteristic. But I don't know what I'd class that with. I don't know how the responses were coming back. [...] It wasn't just copying me, you know. There was enough variation to maintain a level of interest that I would find in an improvisation [with humans].

In a follow-up question, I asked Quinn to elaborate on the idea of a "level of interest":

> **Quinn:** I mean, I was engaged. There was a couple of bits where I had a sort of- she gave me a prod, you know, [*Odessa*], where I kind of upped it a notch or something and that was- that's what you get from playing in improvi– with improvisors. You get a- it pushes you forward a bit, you know, gives you a little prod sometimes. And she gave me a few prods actually, and I don't know how that works.
>
> I wasn't thinking about the sort of formal aspects of where the piece was going. I was kind of chucking it out and sort of trying to find a way to get closer, in the same way that you get closer with a human player. And there was- there seemed to be a learning process going on between both of us. I was learning and [she] was learning, as it were, somehow, and, as I say,

> there was a bit of prodding going on. There was that kind of- [sigh] I don't know how to- what you'd say, apart from prodding really, a stimulus in both directions, I felt. So that was- that kept it lively. It wasn't just churning out by rote.

The above continues to describe the active sense of *Odessa*'s influence, referred to here as a "prod". From the perspective of the system design, this indicates a successful outcome of the Subsumption *Diverge* layer for this performer, which is discussed later on.

Quinn offers several more comments about the notion of a "prod", such as this response:

> ***Was that [previously discussed] engaging feeling [experienced] throughout, or was it more off and on, or were there specific moments that were especially engaging or especially not engaging?***
>
> **Quinn:** Well, yeah. Well, that's where the prods come in really, I mean, I've got my [pause] playing antennae up, you know, the improvising ears are there, and you get into a sort of zone to do stuff, [...] listening to what [she] was doing. And I slowly got more [pause] engaged with it. And then there was the prod aspect, where I thought, well, she's giving me a bit of a shove there, you know, and I felt sometimes I gave her a bit of a shove back.

(The "zone" referred to above is further discussed in the next section.) Continuing with the topic of *Odessa*'s active influence, this response was also given by Quinn:

> ***The next question is if you felt, either throughout or at varying times, if you felt like the complete musical result was part of a mutual collaborative effort, or were there times where you felt like a soloist with an accompanist, or vice versa, or how would you describe that?***
>
> **Quinn:** What was coming back to me was these wonderful prods every now and then, 'cause it was doing its own stuff and sending it back into the mix. So in that- it was really collaborative in that in that sense, I think.

There is an issue raised here about how one perceives another performer, in this case, *Odessa*, and how this perception relates to the establishment of a context for joint activity. The sense given by Quinn is that a potentially contrarian musical response is received positively, under the assumption that the duet partner might make the collaboration more challenging in order to enhance it, rather than to detract from it.

5.2.1.1 The "zone"

When Quinn describes the "zone" an improvisor gets into (above), it relates to the point in the previous section, that (3) *being completely engaged when improvising results in a favourably-experienced improvisation.* For Quinn, there is the suggestion that *Odessa* has facilitated the entering into this zone, which could be described as a mental or psychological state. This relates to a comment by Hamilton, who refers to "the sort of psychological space that, as a player, you're in, when you're actually playing with it [*Odessa*]". On the same topic, after one of the improvisations with *Odessa*, Morgan inquired as to how long it was — a question that also arose for other participants. Morgan justified the question by noting that "I figured I was going a little long, getting in the zone".

This zone that is referred to seems to relate to the concept of 'flow'. The concept of 'flow', initially identified and developed by Mihály Csíkszentmihályi in the 1970s, is especially relevant here in the context of his study of creativity (Csíkszentmihályi 2013 [1996]). Discovered on the basis of qualitative semi-structured interviews about the nature of experience, 'flow' describes the particular experience during the enjoyment of an activity. In his research, he found that "chess players, rock climbers, dancers, and composers," and others who enjoy what they do, were motivated by a "quality of experience they felt when they were involved in the activity. [...] This feeling [...] often involved painful, risky, difficult activities that stretched the person's capacity and involved an element of novelty and discovery" (Csíkszentmihályi 2013 [1996], p. 110). Being "in flow" describes being engaged by and absorbed in an activity, which he proposes is what in fact makes it enjoyable.

For Csíkszentmihályi, flow is linked with the notion of being "in the zone"; although the phrase is more common in sports, it is also the main title of an auto-ethnographic perspective on the flow experience in collaborative piano accompaniment (Brown 2011). Other research has also considered a link between flow and musical collaboration. Byrne et al. (2003) explores flow in collaborative composition through an empirical study in music education. And Seddon (2005), in an empirical study of communication in jazz performance, suggests that further research should investigate the "collaborative 'flow' experience" in jazz improvisation.

In his research on flow and creativity, Csíkszentmihályi (2013 [1996], pp. 111ff.) identifies nine main indicators related to flow that recur repeatedly in the participant data from his extensive qualitative studies. Several of these indicators seem to fit with the experience of improvisation in general, as well as the specific case of improvising with *Odessa*. One of the nine indicators describing the flow experience is that "there is a balance between challenges and skills". Like Wiener's description of an artificial chess player described in the previous chapter, Csíkszentmihályi states that "playing tennis or chess against a much better opponent leads to frustration; against a much weaker opponent, to boredom". Thus, a "really enjoyable game" strikes a balance.

Another relevant indicator describing flow is that "self-consciousness disappears," in the way an "athlete moves at one with the team". This suggests a similar sense to the way a flow state might be experienced in a collaborative musical improvisation. There is also the idea that "the sense of time becomes distorted", that is, expanded or contracted, which reflects the participant experiences with *Odessa*, when they wondered how long they had played for. Additional indicators (of the nine) seem to be relevant as well, but these examples will suffice to suggest that the 'zone' experienced by participants in this study is indeed akin to 'flow'.

It seems that *Odessa* is at times able to facilitate entering into this zone, while there is also a sense that the human performer must play a role in contributing to the preconditions. The human role, in this case, has to do with the how a collaborative performance is approached, from the perspective of an experienced practitioner. As Quinn states:

> (**Quinn:**) [A] kind of closeness developed [in playing with *Odessa*] and in that way you sort of- [pause] you do feel a bit of a relationship with what's going on because, you know, it's developing along with you. So it becomes part of your life [laughs] or your being, you know, your life experience. So it's to do with that level of engagement, and that level of engagement developing through what you're doing. And I think that- [pause] that leads you to feel that it's kind of a more human contact. [...] I wasn't making any concessions, anyway, to it being a machine, in fact.
>
> ***Could you say a little bit more about that, about not making any concessions?***
>
> **Quinn:** 'Cause although that's part of part of improvising, that you're reacting to stimuli and you're providing stimuli, in another level, you're doing a very sort of human thing of sharing something together. And I approach playing [with *Odessa*] in that way, 'cause that's how I approach [collaborative] improvising [in general], I think.

So there is a complex situation in which both co-performers play a role in facilitating the collaborative performance. This relates to the discussion in earlier chapters about the duet partners being tightly coupled subsystems in the collaborative activity. It appears that the human expertise provides a psychological scaffolding that enables the collaboration to take place, but that, on the other hand, the co-performer, *Odessa* in this case, could potentially disrupt the collaborative sense, or prevent it entirely.

In thinking about how one gets into the zone described above, Quinn, in the passage that introduced the term, mentions listening closely ("I've got my playing antennae up, you know, the improvising ears are there [...]"), and also describes a broader approach concerning how co-performers are perceived as partners in collaborative improvisation (giving and getting a "shove"; "these wonderful prods"). Anderson considers similar points in terms of the idea of strategy:

> **Anderson:** It's the same with, you know, regular people, just regular musicians too, just because it's a machine [does not matter]. [...] It's like finding your strategy- I mean, it's kind of funny, because improvisation is also about strategy. It's not just like, 'oh, I hear something, I play', but it's more- also, it's about, like, 'okay, how do I make it work?', like, 'what do I do on top', because you always have these choices. And sometimes you don't think too much about it, and it's kind of like your ears guide you.

Anderson's idea that "sometimes you don't think too much about it" also seems to describe being in the zone, which Anderson connects to the idea that "your ears guide you". This is similar to Quinn's comment (cited above) that "[...] the improvising ears are there and you get into a sort of zone to do stuff".

Morgan's comments (below) also describe similar ideas about entering into a zone, referring to the role of "not thinking about it", and "allowing the sound to take you forward":

> **Morgan:** I wasn't thinking about. I try not to judge- [pause] just trying to put yourself in the place of not thinking about it, while you're doing it, too much.
>
> ***And did [*Odessa*] facilitate that mode for you, of helping you not think about it while you were doing it, or did you have to-***
>
> I had to ignore it. I think it was important to ignore it, you know, in a way, and allow it to come through the subconscious. I mean, to try to play opposite, or upside-down, or all these kind of things, based on what you're hearing, it's actually a little bit too much brain in the process. [...] When

> improvising, you wanna kinda be in that other place, you know, allowing the sound to take you forward. [...] I was basically ignoring him [*Odessa*] to turn off my conscious mind. [...]
>
> ***And do you take that same approach with a person?***
>
> Yes, that's where I go.

However, Morgan did not feel that *Odessa* consistently facilitated entering the zone, but rather, only did so at times. As Morgan put it: "you wanted to [...] escape that- from feeling self-conscious, you know? I think I succeeded a few times".

Walker also identified the notion that listening can allow the collaborative musical output to be a driving force in the improvisation. Like the above sense of letting "your ears guide you" (Anderson) or "allowing the sound to take you forward" (Morgan), Walker described "letting the improvisation lead you to new ideas". However, like Morgan, Walker also found that *Odessa* did not consistently facilitate entering the zone. Following the third improvisation, Walker stated that:

> (**Walker:**) I think this time I realised that I picked up more of the pitches of the program, and those parts of the improvisations were the strongest for me. And that felt that there seemed to be more of a duo partner, when there are more clear pitches involved in the improvisation, or, yeah, fragments of, you know, melodies or whatever, I mean, or pitches [rather] than non-pitched, or more- with the clear pitches, I found that more interesting, these bits, in terms of having feedback, more feedback, and letting the improvisation lead you to new ideas, or to ideas you would do together- develop together with the other, instantaneously.

So for several of the participants, even if only at limited times, *Odessa* does seem able to contribute to the sense of mutual collaboration. (In a few cases, the performance and discussion of the third improvisation indicated that the participant learned something about the system from the previous performances, and that they might, in future performances, choose to focus on what they found had worked best thus far. It is not clear, however, that increased familiarity with the system would increase the sense of mutual collaboration.)

Anderson offers additional support to the point that *Odessa* contributes to the collaboration. While recognising the role of the human performer's approach, and the role of listening, *Odessa* is still given some credit:

> **Anderson:** In the beginning I kind of jumped into it, 'cause I had no idea what to expect, but it stayed interesting throughout.

Another comment by Anderson offers further reinforcement:

> (**Anderson:**) I did find lots of things to take from the partner's playing, and to incorporate into what I am doing. What I mean is, I let it guide the improvisation. And- but it didn't become boring for some reason.

The insistence that it stayed "interesting" and did not become "boring" lends more weight to the role of *Odessa*, even though there is the clear sense that the human performer is substantially contributing to the functioning of the collaboration. The role of the human performer's contribution relates back to Campbell's point that "I felt like I was bending towards the computer's language more than the computer was bending towards my language".

Continuing with the idea of feeling a sense of collaboration, while also recognising that one is facilitating the collaboration with certain actions, Anderson also refers to the idea of "instinct" as an improvisor, which, in this context, is another way of describing the expertise that comes with experience:

> **Anderson:** In a way it's kind of easy for me to play with. [long pause] I don't know, because, I mean, I guess it's like, that- when me, as a- somebody who improvises, that my instinct kicks in. Because even though I know that I'm playing with a computer- but it's- I'm not here just to react, but also trying to make something out of it. So, overall, I mean, I find it interesting, not dry.

The sense of "trying to make something out of it" when performing with *Odessa* seemed to be a common experience for several participants, but they generally do not seem to regard this as substantially different from what one would do with a human co-performer. It does, however, appear that *Odessa* could not consistently facilitate entering into the zone; if it could improvise at the level of a human expert, it could likely be more consistent in this respect. To summarise, the above section, *Odessa* proved capable of facilitating entry into the zone, although for several participants, it was reached only at limited times. For Johnston, discussed in the next section, it was not reached at all.

5.2.2 Lack of responsiveness

Johnston experienced the least successful improvisation with *Odessa*. There was no sense that *Odessa* was interacting, listening, or responding, such that, for Johnston, the entire effort of "trying to make something out of it" was up to the human performer. Ultimately, *Odessa* did not facilitate any experience of a collaborative improvisation for Johnston:

> (**Johnston:**) You didn't really feel you had this ebb and flow interaction.
>
> ***And so how would you describe [*****Odessa*****'s] capability to listen or respond?***
>
> **Johnston:** I would say it was not really listening. [pause] It makes the other per- it makes the person listen how- try to make sense of it. So you, from the start, you may think you're initiating something, but you end up being more passive, so you can try and make it look like its playing with you.
>
> ***Could you say a little bit more about that passive idea?***
>
> **Johnston:** Yeah, so, for instance, once I had realised it's not really- I'm not really interacting with it, then, I think, I'll try and make some music. So let's hear the passages that it's working with, and then work within those constraints. I didn't feel that I could excha- I could give an idea, and that it would be understood or received. So I felt the best thing to do in the last piece was to play a lot slower, and then when it- when I heard a phrase- and then make a phrase similar to what I heard my partner playing.

There were other statements by Johnston to this effect:

> (**Johnston:**) "I felt I had to be more passive. I had to give up a lot more, to try and make something work";
>
> "I thought the best thing to do was to try and use what was there and make it work"; and,
>
> "I felt- I did feel that I had to restrict my vocabulary, to try and make it work".

For Johnston, there is a clear sense that *Odessa* did not provide the right kind of musical interactions to facilitate a collaborative performance.

Continuing with Johnston's unsatisfying experience with *Odessa*, we can note some points of comparison in other participants' comments, even though they had a more satisfying experience. For example, earlier on, Anderson noted that *Odessa* seemed to be collaborating in such a way that "it doesn't become like two people playing almost in a separate room", a contrasting view to Johnston, as shown below. And Morgan described a positive sense of "ignoring" the system at a conscious level, in order to facilitate the entering into a zone where collaborative improvisation can happen. For Johnston, however, ignoring *Odessa* became a negative consequence rather than a positive strategy, and the feeling of playing in separate rooms was, in fact, experienced, rather than avoided:

> **Johnston:** I think, the troub- the thing that it doesn't- the feeling that you're not- that you're actually playing in separate rooms- playing alone, you're playing alo- — usually in a duo with a [human improvisor], even if you've got two different styles, you're trying to find some common ground. And I suppose that's the expectation you have in improvisation, is to try and find common ground. And, after awhile [with *Odessa*], you realise, there just doesn't seem to be any common ground with the person you're playing with. So you end up looking- trying to entertain yourself actually, and just pretending that the other person isn't there, which isn't good for improvisation.

So Johnston ends up ignoring *Odessa* in a way that does not lead to a collaboration. In some respect, Johnston also feels ignored by *Odessa*:

> **Johnston:** I would say I'm playing catchup at the moment. Yeah, I feel like I'm playing catchup. I feel like he's not listening to me. I'm trying to listen to him, but he's not listening to me.

This connects to Stewart's earlier point that there is "not an adequate reward for the attention that I'm paying to it. It just grinds on with its own agenda". However, Stewart also had some positive experiences of collaboration, as did other participants. The positive experiences of collaboration are described in the next section.

5.2.3 Collaboration

In addition to the positive comments already presented, Stewart also states of *Odessa*:

> **Stewart:** He finished off one phrase of mine with a very well chosen note, an ascending phrase. I left the top note out, but I can't remember exactly what the arpeggio was. I think he played a note in exactly the right place. I thought that was smart.

So in considering the many different responses together (including those presented in previous sections), there is a complex relationship between listening, responding, and collaboratively working in parallel.

Morgan, who made the comment about consciously "ignoring" *Odessa* in order to enter into a non-consciously directed collaborative mode, states of *Odessa* that:

> **Morgan:** I felt that we were kind of on the same page, there was like definitely a call and response going on. We were trying to play the same piece somehow.

And Quinn describes a more general sense of the experience playing with *Odessa*: "it really felt like sort of playing, playing a duo". Quinn also notes that:

> (**Quinn:**) So it's- it's- [she] [pause] has found the licks [the phrases that I played], and works with them, and knows how to send them back.

Walker draws an explicit contrast between playing with *Odessa* and playing with an entirely non-responsive source of sound, such as pre-recorded playback[1]:

> ***What could you say about how you- whether or not you felt the system was listening to you, responding to you, or ignoring you? And how did you feel like it was aware of you, in that sense, or aware of your personality, in a sense?***
>
> **Walker:** Yeah, I think it was listening. That I found surprising. That it was really like- I could create some music with the program, in terms of that it would follow me, or do small phrases according to the phrasing I did. And this, I found very interesting. [...] Yeah, it felt like, in this way, I felt it was listening to me. It was also fun to do this together with the program. [...] Yeah. I think it is listening and [pause] it doesn't feel like, you know, doing something with, let's say, the other extreme would be some playback, or something pre-shaped.

[1] In postwar classical music, there are many instances of compositions that use live performers with pre-recorded playback.

This suggests that improvising with *Odessa* lies at a midpoint between collaborating with a human expert and performing with static playback, the latter of which has some similarity to an improvisor who ignores their co-performer, whether due to lack of competence or to preference. Campbell has been cited above several times on this point, but the point is worth repeating here, with some additional context:

> ***Did you feel that, out of the three performances, did you feel that there was a mutual collaborative musical result, or did it feel more like a soloist / accompaniment scenario, or how would you describe that?***
>
> **Campbell:** I would say, I mean, it was definitely a duo, but I felt like I was [pause] bending towards the computer's language more than the computer was bending towards my language, which is fine. You have to do that sometimes in performance [with human musicians] too.

5.2.3.1 Leading and following

The complex relationship between listening, responding, and collaboratively working in parallel has an additional component in the issue of leading and following. Although neither term is explicitly raised here, there is an implicit issue of leading and following when Campbell states that:

> (**Campbell:**) I was being influenced by it [because] I was trying to play with it. So, there were times that maybe I would have done something else, but I ended up doing something different.

The issue of following did, however, explicitly arise with some of the other participants, such as Walker:

> ***You mentioned that you noticed that it would follow you sometimes. Did you ever feel that you would follow it?***
>
> **Walker:** Yeah, I felt that. When- I guess when I mentioned, you know, [that] it didn't work so well for me in terms of finding timbres together, but more in terms of more pitch-related sounds, in that way I was following the program. 'Cause I think, in a way, I looked for something — 'what can I do with the program?' — which is interesting to me. So, in that way, I- yeah, I guess I followed it.

Anderson seems to recognise the complexity of this topic mid-speech:

> **Anderson:** I think in general I felt like I was following it, but not in a way — how can I say? — [long pause] interesting. I was following it, but there were certain points where I was kind of waiting for it. I mean, that's kind of interesting because waiting for it means I'm not really following it. But, it didn't feel like it was following me. But then, also, it didn't feel like- that I was chasing after it. Does that make sense? I think there's a difference between following and chasing.

While Anderson did not feel *Odessa* was following, Morgan had the opposite experience with *Odessa*, namely, that it was doing too much following:

> **Morgan:** I felt the- I wanted him- her- to [pause] like, [I] was trying to shake a tail. [It's] like I was being tailed, you know what I'm saying? You know? It's hard to escape. Is that a fair assumption? [pause] But not all the time.

In thinking and speaking about this topic at length, Anderson came to a view of the collaboration with *Odessa* that seems to strike a balance between leading and following:

> **Anderson:** I cannot say [...] it was in control, you know. I mean, I could just say, "well, that was my choice" [to let it be in control]; [if so,] that means I was in control. But it's never that clear. [...] It's kind of like give and take. It's something that circulates between two voices- persons- musicians- it and me.

It seems that *Odessa* is capable of allowing this on-going "give and take", as long as the human co-performer can find a way of relating to *Odessa*'s musical behaviour.

5.2.3.2 Partnership

It is significant that Walker, below, draws a parallel between duo collaborations with human improvisors, and those with *Odessa*; this parallel is drawn on the basis that, in any such duo, both co-performers seem to reciprocally affect one another:

> **Walker:** I guess in the third set I probably- I could find the- with the program- the point where we could meet as partners, so to speak — you

> know what I said about this pitch-related [point of connection with *Odessa*]? That's also different, I mean, with every musician you play with. Somebody- I mean, like, if you would compare the duo partners — my duo partners — although it's always me [improvising], it- I- the music is somehow quite different, depending on the other person's personality, and on the music we just- we build together.

The idea that "we meet as partners", in the sense that *Odessa* is a legitimate co-performer, capable of mutual collaboration, also arose for Anderson:

> ***Did you feel like you were co-creating the music together?***
>
> **Anderson:** Interestingly enough, yeah.
>
> ***And so [...] would [it] be fair to say that [...] there was a sense in which you were playing* with *each other?***
>
> **Anderson:** Yeah, interestingly, yeah, because, it's like, logically- I mean, in theory, I know it's not really *with* me, but [long pause] but, I mean, yeah- but the thing is, yeah- [...] I mean, it made sense to me musically.

Later on in the interview with Anderson, this notion of being "with" each other, in the sense of the partner having a presence in a shared situation, is taken up again:

> **Anderson:** You've seen lots of cases that people- sometimes, [even if] they're good, some of the time, they're just "not there". And why is that? Can you just- how can you say to them, that- how can you sum up that person? It's the same thing. And also, a good improvisor is not just somebody who can respond. I mean, it's just something that's more complex than this just one thing to say, so I don't know. [...]
>
> ***So, in this case, if you had to think about the three improvisations [you did with* Odessa*], how would you describe it being "there" or "not there"?***
>
> **Anderson:** If I just totally just get rid of this notion that it's "it", then I will say it's there.

Note that the issue of responsiveness is raised, but contextualised within the idea that collaborative improvisation is not easily reduced to the notion of direct responsiveness.

5.2.3.3 "Strategy" or "approach"

The topics being discussed here — listening, responding, collaboratively working in parallel, leading, following, and being present or "with" someone as a legitimate co-performer capable of mutual collaboration — are, for some participants, directly connected to their approach or strategy. Earlier on, Anderson was cited stating that "improvisation is also about strategy". As shown in this section, some participants reflected on how their collaboration strategy related to their experience and interpretation of *Odessa*'s actions. Take, for example, Hamilton's account:

> **Hamilton:** From my end of things, I was trying to draw out different material from the- my duet partner, by doing things that really had nothing to do with [*Odessa*'s] piano at all, especially non-pitched materials. So it's hard to know if- what effect it had, because there were these long silences. But the effect it created was- I thought was interesting. It almost was this kind of trading back and forth of ideas between the partner and me.

So the idea of "trying to draw out different material" from *Odessa* as a strategy led to an experience of what seemed close to the "trading back and forth of ideas". It is interesting to note that this contrasts with Walker's experience (recounted above), who preferred working with *Odessa* on pitched rather than non-pitched materials. Hamilton used a different strategy in another improvisation for the study:

> **Hamilton:** That one was really different [from the previous two improvisations]. My tack for that one was to see what would happen if I just sort of fed it continuous material within a really limited pitch range. [...] The way it reacted was super predictable. It almost behaved like a, I dunno, like a well-behaved improvisor. I mean, it kind of mirrored what I did, but then was adding material. When I would change to a new collection of a pitches, it would also move chromatically down. I mean, I started with I think F and E and D, and then, when I moved to Eb, D, and C, it almost immediately recognised the change. Likewise, when I went to the high register. It was almost too much ghosting — following around — which I was really surprised at.
>
> So I kind of started to test it, and I was even more surprised when I dropped out and, all of a sudden, it switched down to the lower register and immediately started playing a different kind of material, which, kind of, as a structural decision, was actually pretty cool. [That's] something you

> would sort of hear from a sort of experienced improvisor, human, who can just, on a dime, switch the texture up to create a different sense of the form.

The surprise Hamilton experienced seems again like a positive design result of the Subsumption *Diverge* layer, while the idea of "almost too much ghosting" or "following around" seems to be a drawback of the Subsumption *Adapt* layer. The latter issue was also commented on by Morgan:

> (**Morgan:**) It's eerie, you know? It's sort of [like me] getting parroted [by *Odessa*], you know? Altogether, it wasn't so bad. [pause] [...] It sort of gets what you're doing very accurately — you know? — which is unnerving. Almost a little too much, you know what I mean? You don't really wanna be shadowed that much.

Regarding strategy, Hamilton's notion of testing it ("I kind of started to test it") also arises for Quinn: "Part of me was sort of trying to test out the playing partner"; and for Stewart, who notes a similarity to performing with a human musician:

> **Stewart:** It's also true that when I play with somebody new, I test the system as it were, or test *their* system. So it's not completely unlike playing with somebody new. But you wouldn't be as ruthless about testing the system with a human being, I don't think, for fear of hurting their sensibilities.

This suggests that for Stewart, and, to an extent, for others, *Odessa* was not being approached in the exactly same way one would a human improvisor, yet the approach was also not entirely different. The potential to positively or negatively affect another person in a collaborative improvisation supports Lewis' (1999) point that "interaction and behaviour are carriers for a complex symbolic signal, and that notes, timbres, melodies, durations, and other music-theoretical constructs are not ends in themselves. Embedded in these objects is a more complex, indirect, powerful signal" (p. 106), one that reflects the social character of musical interaction.

The next section considers all of the above sections in this chapter in order to make an assessment of *Odessa*'s design on the basis of the participant experiences documented in the study.

5.3 Design assessment

This section focuses on assessing the design of *Odessa* in terms of the research questions, which are concerned with its apparent intentionality, its capacity to collaborate and be engaging, and the relation of these concerns to the system architecture. To facilitate this discussion, I briefly introduce one final theme that will also serve as a transition directly into the assessment. Though not introduced above, this theme recurs at various points in the comments given throughout the chapter, namely, the way the participants refer to *Odessa*.

Most participants have typically used a mixture of ways of referring to *Odessa*, although some consistently use only a few terms of reference. There is a range of reference terms that could be reasonably expected, from the "duet partner", a term suggested by its use in the instructions for the experiment, to the "program", the "device", the "computer", and similar, and simply "it", which I also used in the interview questions. More interesting, however, is the occasional use of "he/his/him", "she/her", and what is referred to in language studies as the "singular they", also called the "epicene they", meaning it does not denote either male or female (Balhorn 2004; Sklar 1988). In some cases, participants verbally struggle before settling on "it", "he", or "she", whereas "they", referring only to *Odessa*, often seems to arise more naturally. The linguistic explanation for this is that when an "animate/human substantive" has a referent with an indefinite gender, it is common usage to use a plural pronoun ("they/their/them", rather than "it/its") for agreement (Sklar 1988).

So *Odessa* is frequently regarded as animate by the participants interacting with it, while, at the same time, it is not perceived as obviously male or female, although some participants settle on a choice between male and female early on. This suggests, with respect to the discussion in previous chapters, that the cues for agency evoked by the Subsumption and interaction design are at least partly effective. These cues are in part provided by *Odessa*'s lowest layer, *Play*, but also by the alternating activation of multiple layers.

The highest layer, *Diverge*, was also at least partly successful, in that several participants felt that they were being challenged (prodded, etc.), to the benefit of the collaboration. This later also played a role in the perception that the system was collaborating in a parallel way, neither simply leading nor simply following or responding, which had to do with the interplay between the *Adapt* and *Diverge* layers. This interplay also further supported the perception that *Odessa* has intentional agency, meaning that it was perceived as having its own approach, logic, or language, that it applied to the collaborative effort, even if it was also perceived as having limited competency.

The limited competency of *Odessa* is further reflected in a statement by Hamilton, but the underlying effectiveness of the design is also highlighted: "there's all kinds of things that it's not doing, but, within its own parameters, it seems to be pretty flexible, which is kind of what you look for with another player". Stewart also made additional comments, not covered above, that seem to suggest a sense of intentional agency and flexibility, even while referring to *Odessa* as "it": "what's good about it [*Odessa*] is, it doesn't seem to be upset by anything I'm doing, which, again, is not always the case [with human players]". Stewart re-iterates the point later on, along with an additional related point: "I've said that the great thing about it was it didn't seem to be upset by anything that I was doing, and it didn't seem to be bored by it, not as far as I can tell". Clearly, the system behaviour facilitated these assessments, which point to the strengths of the design.

A few of the main drawbacks of the system design seemed to concern the *Adapt* layer. When *Odessa* followed the participant's musical activity, it followed too closely. This suggests that the transfer of input pitches to output pitches, which typically takes close to one tenth of one second, is too immediate. Also, *Odessa* did not seem to allow for extended musical sections of similar material that would facilitate a sustained musical exploration or development. This was likely due to two issues. The first is that small gaps between input pitches could be frequent, due to the exhaustion of available input for output, which often activates the arbitrary pitch generation. The second is that, even if a musical segment in the collaborative performance was sustained, it was too easily perceived as not developing or sounding childlike, due to a phenomenon akin to a feedback loop: if a small collection of pitches was played by the participant, then, on that basis, played by *Odessa*, then, in turn, played by the participant again, the human participant likely begins to aim for a more complex process (e.g., deliberately exploring slight variations in a quasi-repeating note pattern, or perhaps working toward a palpably physiological listening response, even with a precisely repeating pattern). In this situation, however, *Odessa* simply remains locked into a quasi-repeating note pattern with merely arbitrary variation, which does not suggest sustained exploration or development.

The next chapter presents my solution to this problem, and the results of a follow-up study with the second iteration of *Odessa* that was based on this solution.

Chapter 6

Follow-up study

As stated at the close of the previous chapter, several of the main drawbacks of *Odessa*'s initial design relate to the *Adapt* layer, in particular, to the near-immediate translation from input to output pitches (typically less than one tenth of one second). This led to either rapid changes in sets of output pitches, or the system remaining on a limited set of output pitches played in a largely repetitive pattern, as described as the close of the previous chapter. As a way to overcome both of these problems, I sought to incorporate something akin to a pitch "memory". However, I wanted to remain faithful to both the Subsumption approach under investigation — namely, the interaction of simple mechanisms used to produce complex behaviour — and also remain theoretically aligned with the ecological view of perception that led to the use of Subsumption in the first place, and that has proved effective in theorising the cognitive aspects of free improvisation.

6.1 Theoretical background for extending the design

In their co-edited volume on dynamical systems and cognition, Robert Port and Timothy van Gelder state that "the dynamical approach to cognition has turned out to be deeply compatible (though by no means identical) with Gibsonian or ecological psychology" (Port and Van Gelder 1995, p. 373). The authors lend their support to the argument that dynamical systems theory has a critical role to play in providing mathematical analyses of cognitive phenomena identified with the ecological approach to perception and action. Although the mathematical analysis of cognitive phenomena is beyond the scope of the present study, it is of interest that one of the chapters in their edited volume is on human audition (Port, Cummins, and McAuley 1995).

In the above-mentioned chapter, Port, Cummins, and McAuley (1995) criticise the idea that we can take what they term “scientific time” — namely, the subdivision of time into standardised units (e.g., seconds) — and apply it to what they term “biological time”, namely, the way animals (including humans) generally use temporal information in perception and cognition. Their main subject matter is sound and audition, and in this context, they state that “among psychologists, linguists, and phoneticians and [...] probably [...] most laypeople as well,” there is a widespread assumption that we store our perceptions in memory “buffers”, placeholders that contain information with an intrinsic duration, matching that of scientific time (pp. 343–345). In other words, the widespread assumption they describe (and go on to criticise) is that (animal or human) memory is like a tape recorder, such that, for example, a three-minute song heard on the radio would be stored in memory in a three-minute long buffer. This assumption also implies that to access, for example, the second minute of a song, one simply begins at the implicit “two-minute mark” in the segment. They term this the “naive view of time,” which they acknowledge is well justified by our intuition, but argue that it is actually false (pp. 345–346).

It is also beyond the scope of the present study to recount their extensive arguments against the naive view of time, and to offer their concrete alternate proposals, which involve neural networks to facilitate pattern recognition and the adaptation of action to perception (e.g., clapping one’s hands to a musical pulse). For the studies on *Odessa*, I have set out to investigate a more extreme parsimonious model that does not rely on neural networks, however simple. My goal has been to identify what can be achieved using only a basic Subsumption system. However, as one of the chief limitations of *Odessa* was the “closeness” of action to perception in the Subsumption *Adapt* layer, the evidence suggested that a design extension was needed for maintaining adaptive system output without too closely following system input. Given that Port, Cummins, and McAuley’s arguments about dynamical systems and cognition are “deeply compatible” with the ecological approach that has informed my understanding of collaborative improvisation, I sought to find a simple, non-*neural network*-based alternative for *Odessa* that still adhered to their basic premise.

An alternate solution to the problem of memory that was considered for *Odessa*, but ruled out, was a Subsumption-related idea given by Mataric and Brooks (Mataric 1990; Brooks 1999, chap. 3). This approach stored long-term information (during a single continuous operation) about the existence and relative locations of landmarks in a room, using a network of additional layers for storage. The overall context of this approach did not prove relevant here, as it relies on the idea that the same landmark, defined in spatial terms, can be returned to at a later time. With temporal information such as that of music, there are no landmark equivalents that can be returned to in this sense. One could imagine that a repeated melody is in some sense a return to an original melody, but this insight does not aid in dealing with free improvisation, which could potentially never repeat a perceived musical event in a given performance.

6.2 Design extension

Port, Cummins, and McAuley (1995) are concerned with time; I had the additional concern of pitch. These authors discuss entrainment as a general basis of adaptation and pattern recognition. Entrainment has also been discussed in the context of ethnomusicology by Clayton et al. (2005, p. 4), who define it as "the interaction and consequent synchronization of two or more rhythmic processes or oscillators," consistent with the definition in Port, Cummins, and McAuley (1995). A full discussion of entrainment is beyond the present scope, however, as the present study does not use a biomimetic model and does not purport to be an investigation of entrainment as such. Rather, the theory of entrainment (especially given its apparent relation to ecological psychology) has suggested a way to implement a pitch-based memory-like system for *Odessa*.

Specifically, my solution to pitch memory was to add a module to *Odessa* with a virtual oscillator for each discrete pitch, which would get "excited" by incoming pitches (i.e., would entrain to them), and gradually decay. Thus, input to the system is, as before, rapidly taken in, and output is still rapidly produced, but — rather than a direct transfer of input pitch to output pitch, as in the first iteration — for the second iteration, the input pitch is directed to the memory module, and the output is taken from a random selection of still excited frequencies (Figure 6). All pitches have designated independent registers, and all decay independently at an equal rate, returning to a resting state after ten seconds. This duration was chosen after empirical testing, on the basis that it seemed to adapt output well to both gradual and rapid changes in input. If an input pitch is repeated while its equivalent is still excited in memory, then, regardless of where it is in the decay process, the equivalent pitch in the memory is maximally excited again.

While this module, in its current form, precludes pattern recognition, it does seem to offer closely related pitch patterns on the basis of arbitrary combinations of excited pitches. It also prevents the effect of *Odessa* "too closely following", originally found to be problematic by participants. Moreover, it avoids the naive memory buffer model; it remains a simple Subsumption design; and it remains compatible with the ecological analysis of improvisation that led to using a Subsumption design in the first place. On the other hand, the lack of pattern recognition is a significant limitation, for which appropriate solutions will be considered in future research. The relation between *following* and pattern recognition is considered in the discussion at the end of this chapter.

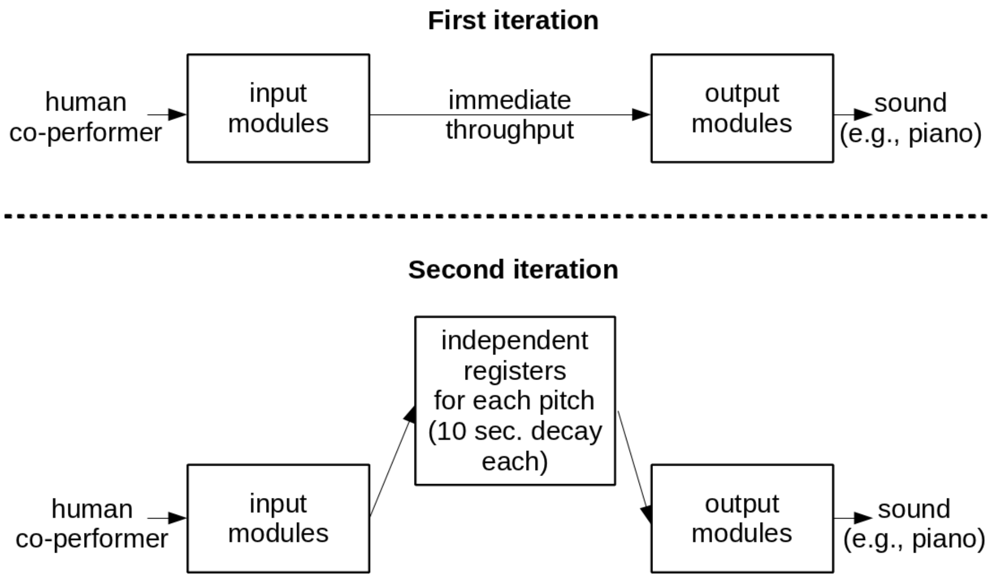


Figure 6: Memory module.

6.3 Methodology of follow-up study

After implementing this memory module, a follow-up study was undertaken with the second iteration of the system. This study departed from the initial study in other ways, as well. For one, a major shortcoming of the first study, the synthesised piano sound, was remedied by using a electromechanically controlled acoustic piano (Yamaha Disklavier) for the follow-up study. This also meant, however, that some points of comparison between studies were expected to relate mainly to the shift from synthesised to acoustic piano.

The follow-up study used two participants from the first study, to facilitate a comparison across studies. This had the drawback that the participants were, at this point, aware of the system design and the general experimental approach and initial results, as they had been sent a paper on the system design and first study (Linson et al. in peer review). The follow-up study also added a trio performance, with both human participants and *Odessa*. For the trio performance, the two instruments played by humans were mixed at equal levels into a composite mono signal, which was received by Odessa as input. The design extension was not disclosed to the participants prior to the study.

Given these circumstances, despite the continuities between the studies, the follow-up study was used to explore themes that were not covered by or influenced by the first study. For example, the second study was not a valid means for finding significance in how *Odessa* was referred to (e.g., as “he”, “she”, or “it”), since this was already explicitly thematised in the previous findings, which had been disclosed to the participants. The focus on different themes made it possible to allow the follow-up study participants’ identities to be disclosed, with their explicit permission, without revealing any identities in the first study. The follow-up study participants, Evan Parker and John Russell, have also kindly granted their permission to use the musical audio from the follow-up study as supplementary material. This access will, among other benefits, allow the reader to observe how *Odessa* functions with different instruments, in this case, soprano saxophone and acoustic guitar, respectively.

After each of their duet performances with *Odessa*, they were asked to speak about their experience in a semi-structured interview that followed an unstructured think-aloud protocol, as described in the first study. They were not permitted to hear each other's verbal responses for either of their duet performances. Before and after the trio performance, however, I conducted a group interview. The sequence of events was as follows:

(A) participant 1 performance;

(B) participant 1 private discussion of (A);

(C) participant 2 performance;

(D) participant 2 private discussion (C);

(E) participants 1 and 2, group discussion of (A) – (D);

(F) participants 1 and 2, group performance (trio); and

(G) participants 1 and 2, group discussion of (F).

During (E), that is, just prior to the trio performance, I asked the participants to jointly consider some points they had each made after their duet performances, points that seemed to be related (which I repeated back to them at this point). In the group interview, I assured them that their views did not need to coincide, and that divergent assessments did not need to be altered. This allowed me to discover if points they made were indeed related, or if they in fact reflected differing assessments. After the trio performance, they were interviewed about it together (G). Again, differing assessments were explicitly welcomed, as well as points of agreement, in order to try to discover to what extent their experiences were similar to one another, if at all.

6.4 Results of follow-up study

The follow-up study primarily addresses whether or not the additional module played a significant role in the experience of *Odessa*'s behaviour. To a more limited extent, the follow-up study ascertains if any results of the first study may have been affected by the use of synthesised input. All comments below are taken from the individual sessions, until the group interview is explicitly introduced.

The most direct evidence of a change brought about by the additional module that affected the participant experience with *Odessa* is given by Russell:

> **JR:** There was that phrase, which wasn't immediate, and it was from something I played some time before it came back at me. But that could have just come from my playing anyway, so I don't know whether it comes from the machine or not. So I don't know how much is kind of an immediate thing, or, yeah, it's got to be. It's not kind of responding to something 30 seconds ago and sticking that in. But it felt like that sometimes, just in the way it worked. So it felt like there was a kind of broader identity than just a sort of immediate kind of stimulus–response thing in the present, basically.

This account implicitly describes what I regard as an improvement on the previous iteration, where the immediacy of the input–output transfer was a source of problems for several participants in the first study. A similar point also seems to be noticed, although less explicitly, elsewhere in the interview with Russell:

> ***Would you say there was a back and forth [between you and* Odessa*]?***
>
> **JR:** I think there most definitely is, actually, and I'm not sure where the back-and-forthness is coming [from], 'cause I felt some of my stuff [was played back to me], [though] I can't remember I played it, but it was one of my licks. [...] I don't know where that came from in terms of the linearity of the piece, really. But, you know, I was being quite cavalier with the pitches [I was using, playing] around a bit, and the program was spot on in getting them. [...] I think some of it came from the program, you know.

The above describes what seems to be a direct effect of the excitation of multiple pitches over time in the new module, and in the way that the pitches are arbitrarily selected and re-introduced as system output. I also regard this as in improvement over the previous iteration.

The remaining evidence of the impact of the additional module is more indirect. This is because I anticipated a level of system performance on the basis of the design premises that was not met with the first iteration; this suggested to me that I had not reached a limitation of the design premises, but of the implementation. The first study provided feedback about the system's limitations, some of which were addressed in the second prototype while still adhering to the original design premises. While the implementation could, in principle, continue to be improved, the follow-up study with the second iteration produced substantial feedback about precisely what I anticipated as the potential limitations of the design premises. To overcome the problems raised in the follow-up study, it would seem that design components would be called for that would exceed the restrictions of a system design that could be appropriately regarded as a Subsumption system, at least as far as system complexity is concerned.

From this perspective, the additional module can be viewed as having an impact on the performer comments that, even when critical, appeared to be more relevant to the design than the implementation. The following are two key examples of what I view as being more relevant to the design. The interaction with the second prototype appeared to elicit these comments, which did not seem to occur in this form in the previous study. In the first example, Parker states:

> **EP:** Dynamics are also perhaps an issue. It doesn't- [...] What would you call it? The emotional- those kind of dynamics. What do dynamics mean normally? They have some kind of emotional significance, perhaps, or dramaturgical connotations. It doesn't have much sense of dramaturgy either.

The recognition of emotional significance in music, and an understanding of musical dramaturgy suggest the need for a more elaborate collection of mechanisms, with a greater degree of internal system complexity.

In the second example of feedback that I consider especially relevant to the design limitations, Russell states:

> **JR:** It wasn't dealing with the pitches in the way that you kind of expect it to deal with the pitches, if it had played in a Blues band in the past [...], so it wasn't like it was coming from a position of knowing the Blues and deconstructing it; it was coming from the position that it was just working with what it had got, that it was sort of deconstructed anyway, [...] you know. So it can't play the- [...] can't quite get inside the humourous- [...] It's not got a sense of irony that a human would have for the same, something like that. But of course, I wasn't expecting that.

Again, what is described here is both an understanding of a particular musical form, the Blues, as well as a humorous way of relating to that musical form. The above description suggests that it is not sufficient to simply infer Blues-related musical information from the other performer, as would be expected of a knowledge-free Subsumption system. Rather, what is called for is a specific way of relating to the musical material, such that it demonstrates an understanding of how the material is being manipulated, and to what end. This also suggests the need for a more elaborate collection of mechanisms, with a greater degree of internal system complexity.

6.4.1 Machine understanding

The general problem of *Odessa*'s perceived lack of understanding of what is taking place in a piece, and with its human co-performer, arose with the previous iteration of the system as well. The additional module in the second iteration did not remedy this problem. This is especially clear when Parker states that:

> **EP:** As long as you play on the piano's- or, the device's terms, you can go along quite well. But there are two problems. One is the sense that it doesn't really have, or doesn't feel as though it has, a sense of the total form. So it doesn't really know when an ending might be coming; [...] and, furthermore, [...] it doesn't realise when it has done something which has made a significant impact on me, or where I've taken something from the piano and either developed that, or transposed it, or imitated it. It doesn't seem to recognise any of those standard ways of interacting with another musician.

These are deep musical-structural issues that psychological cues alone cannot address. While cues likely play a role in understanding the total form and recognising a potential ending, these aspects of music presuppose a broader understanding, a background against which such cues would be intelligible, even if unconsciously (see Doffman 2011). Without this background, which is generally absent in any Subsumption system, it cannot succeed in these tasks. There are no *a priori* rules that would identify any ending or any total form in a free improvisation, as there can always be an innovation that subverts any previously established convention in these respects.

As Russell points out, an ending in a free improvisation may arise from factors that are musically intrinsic, but may also involve other judgements related to a broader human experience as well:

> **JR:** There was a bit somewhere where I thought, well there's a thing that it carried on a bit, as I said, it got the last word in. So it doesn't look for

> endings in the same way as a human musician would do. So that was- it's got a bit of incessantness about it that is kind of not human in a way. I mean, I tend to sort of think in terms of, 'oh, how long have we played? oh, that's enough, we'll have an interval now', or, 'oh, that seems like a nice place to stop', as well, you know, in terms of a piece or the shape of something.

That kind of complex judgement about when to end a piece appears to be an example of the "know-how" of expertise that results from extensive embodied experience, rather than being derived from *a priori* rules, as described by Dreyfus (1992 [1972]) and, more recently, by Dreyfus and Dreyfus (2005). Along these lines, a detailed account of improvised musical endings is given in Doffman (2011).

6.4.2 Playfulness

Following up on Russell's comment about *Odessa*'s "incessantness", I asked him if this made it "more like playing along with a recording". He responded:

> **JR:** No, it's not like playing along with a recording at all, because there's a definite feeling that, you know, it's responding to what you're doing. And I know it's not supposed to do that, but it's what I'm kind of putting on the program; that's not actually there, it's what I'm imagining. But it's really responding to what you're doing, which you wouldn't get from a recording. Playing along to a recording is a completely different feel. [Playing with *Odessa* is] like playing with a chum, really, you know, albeit a robot chum. But it's, yeah, you can't improvise without having that sense that, you know, there's something coming back. I mean trying to improvise to pre-recorded stuff is awful.

In the above commentary, based on his knowledge of the first iteration of the system and his experience in the previous study, Russell concedes that he is playing an active role in interpreting the cues for intentional agency. However, he is also reflecting on his experience and suggesting that imagination alone does not typically seem to be sufficient for attributing responsiveness to the other source of musical material. Russell eventually concludes that system is not only responsive, but also playful:

> **JR:** The program is very playful, that's the point I'm trying to make. It's very playful, and that's fun. That's the nitty-gritty of improvising, the act of playing and doing it, and the program responds enough to be playful.

This experience points to one of the strengths of the design. It is significant that even with its design parsimony, *Odessa* is able to engage in a playful collaboration with an expert improvisor.

6.4.3 Impact of acoustic piano

The use of the acoustic piano was an enhancement to the interaction experience with *Odessa* that was welcomed by the participants. While the acoustic piano lent a more typical sonic quality to the performance situation, it did not seem to fundamentally alter the perception of the system behaviour. However, there are two points worth noting about the piano in the follow-up study. The first, pointed out by Parker, is that an acoustic piano has certain sonic properties that can be musically interacted with independently of how it is played. As he describes:

> **EP:** [With a piano,] the hammer hits the string, the soundboard resonates. The piano sounds right, whatever you do, it sounds in tune. [...] Once the fingers hit the key, the instrument does the rest, and it does it rather well. And it resonates, you know. It does all of what happens next, largely without effort.

Parker goes on to explain two parts of the challenge of playing with *Odessa*: “one part is the challenge of playing with any decent piano, or any piano. The other part is to do with the [programmed system behaviour]”. To me, this suggests that an ideal system would not only be able to take account of what the other performer is doing, sonically and behaviourally, but also be able to take account of the real-time sonic properties of its own instrument (in this case, the piano), as well as the way the acoustic properties of the collaborative performance function in the space as a whole.

A different point was raised by Russell, concerning the relationship between instruments in the duet, namely the “acoustic guitar and piano”. Due to this speech construction, one cannot assume he is referring to an acoustic piano, although similar comments did not arise in the previous study with the synthesised piano. Also, it cannot be ruled out that the below comments may be influenced by not only the piano, but also by the room and other experimental conditions. However, the acoustic piano seems to better facilitate a comparison to experiences with other human duet partners who used an acoustic piano, as indicated by Russell’s comments:

> **JR:** It’s very difficult for acoustic guitar and piano. It’s one of the hardest kind of instrumental combinations. [...] The other thing [*Odessa*] did which

> I thought was great, because- what you sometimes get with piano players in the real world- [...] they just follow what the guitar player is doing and then just swamp it, you know, so [...] you go up high, and then they jump on your back and they're doing all this high [playing], you know. So you go low and then they follow you there and then- so you play fast, and they play even faster, 'cause they can do that, it's the nature of the instrument. [...] I was getting well swamped by piano players who just did their stuff and thought they were playing close to you but, in fact, they were swamping what you were doing; there was no transparency. And I find with this program, actually, that although it's got that incessant thing, [...] when you're doing it, it's somehow- there's a transparency in it for the guitar, which is very nice.

The above suggests that the duet performers are indeed functioning as tightly coupled subsystems, in a way that preserves an equilibrium or homeostasis in the collaboration. While this is a positive result of the design goal, there is another negative sense of homeostasis that was identified when the system performance was considered in isolation; this is examined in the next section.

6.4.4 Homeostasis and style

As described above, the piano has acoustic properties that are, to a large extent, fixed, which Parker described at one point as "stubborn". Parker also stated that *Odessa* generally remains within its "comfort zone", or has a "customary mode of response", which to me also suggested a stubborn quality. I asked him about these two senses of "stubborn", a sonic stubbornness and a behavioural stubbornness, to find out to what extent, if any, they were related. Parker responded:

> **EP:** Well, I think it's got both those things. And they're both parts of the challenge of playing with it. One part is the challenge of playing with any decent piano, or any piano. The other part is to do with the [programmed system behaviour].
>
> ***Did you feel that there was a sense of individuality of a player "behind" the piano, or did it seem more that you were just-***
>
> **EP:** Well, in a sense, yes, in a sense. Because it had [...] to do with [...] a kind of restriction in terms of dynamics, restriction in terms of, to a large extent, speed. Those were the two main areas where the device seemed to have a style, if you want to call it that — an approach.

The above-described "restriction" or limitation in *Odessa*'s capacities suggest that it had a narrow range of behaviour. It remained stable, in a homeostasis, but the feedback from the human performer did not facilitate large deviations from its stable state, which could have led to more adventurous improvisations. Its narrow range of behaviour still led to it being perceived as having a personality, although perhaps a not especially adventurous personality.

Russell describes the narrow range of behaviour in terms of "clichés", but also recognises its personality:

> **JR:** I mean, that's how we play as humans. We have our clichés, don't we? And it felt a bit like [*Odessa*] had got some of [its own] clichés. Whether [it] was getting them from me and reworking them through a program that has a different way of generating these clichés- but, there was definite stuff there to kind of grab onto, I think, that had a personality. It wasn't, you know, it didn't feel like it was a kind of random response. You know, there was something under the hood that was- that had some kind of artificial soul or, you know, like personality or something to it. Whether that was- so that would be in terms of- I guess some stylistic things. I felt there was some stylistic things in there that were peculiar to the program actually.

I found it interesting that both Parker and Russell came upon the idea of style at the conclusion of each of the above quotes. Parker states that "the device seemed to have a style, if you want to call it that — an approach"; and Russell concludes with: "[...] I guess some stylistic things. I felt there was some stylistic things in there that were peculiar to the program actually". Both of them were reluctant or hesitant, it seems, to use that terminology, so I thought I might learn something more by asking them about it in a group interview, prior to the trio performance.

6.4.5 Group interview (pre-trio)

The below interview excerpt is more easily understood as a whole, so I will reserve commentary until afterwards. Bold typeface indicates my dialogue, and initials are used to distinguish the speakers. After paraphrasing their comments on style to them, I inquired as to whether or not they might have been making a similar point:

> **JR:** It's really difficult when you talk about style because, you can say style in terms of genre and all of those kinds of things. But then, there's the sort

> of musicality or something, and that's like a different sort of approach to style. [...]
>
> ***Personal style?***
>
> **JR:** Yeah, as well as the instrumentation. [...] It's not just- we've got these instruments in the orchestration, but we've also got these characters, these people. And they have particular things that they do and don't do, or prefer, to others, you know. [...] So style is, I mean, a bit like territory. It's kind of, everybody has their licks, clichés, their territory, the material they're working on at a particular time, [...] their musicality. [...]
>
> **EP:** I don't know if it's got a style, in that sense. [...] In a way, I'm saying I find it hard to distinguish its- the device's limitation from whatever might be its style.
>
> ***I'm wondering if that relates to what you, John [Russell], said about the idea that it kind of has a few areas that it goes to, that- or you even described them as possibly a little bit cliché, as far as its own playing is concerned. That sounds like- I would read that as a limitation.***
>
> **JR:** Yeah, would you call it a sort of default position, or something like that? It seemed to go into this sort of thing.

Based on this interview, it appears they agree about the system's narrow range of behaviour, which is called its "limitation" or "default position". However, Russell sees a relationship between *Odessa*'s behaviour and a personal musical style that a human improvisor might have, whereas, prior to the trio performance, Parker does not view a relationship in this respect.

6.4.6 Trio

The group interview continued after the trio performance, in which Parker had a more favourable experience:

> **EP:** I found it integrated better into the trio, and it was easier to play with the device in the context. Partly because, if you think of a tripod as opposed to a bicycle, you know, the bicycle doesn't really stand up without that extendible leg, unless there's something moving it along. So this had

> some of the stability of, well, let's say a tricycle, because it was moving along. [...] I can think of equivalent things I've done in the so-called "real world", which isn't much more real than this. [...] I quite enjoyed the piano playing in that [trio] context. And, as I say, I found it easier to deal with because I knew John [Russell] would carry the main line for a bit, so I can be over here, sort of playing accompaniment with the piano, and then shift to another– go directly to John [Russell] and see what the piano would do.

It is interesting that, although he does not refer to the notion of personality, there is a sense conveyed that *Odessa* was able to serve different roles, including jointly accompanying Russell with Parker, and that this was able to facilitate an experience more similar to experiences with human players.

I was interested to discover whether the different roles the system played were experienced as discontinuous "jumps", which were occasionally experienced by participants in the first study:

> ***You said you could imagine analogues between this and real world situations. Would it have seemed that there was one person sitting at the piano the whole time, or that someone left and then someone else came along?***
>
> **EP:** No, it didn't feel like that to me, it felt like one person.
>
> **JR:** No, it's the same piano player, as it were, there.

This suggests that the additional module, in the context of the interplay between Subsumption layers, resulted in a more coherent identity for the system.

6.5 Discussion

With the second iteration of the system, given the experimental evidence, it seems reasonable to conclude that some of *Odessa*'s limitations can be viewed as a result of the design premises. The results of the first study led to improvements in the implementation without fundamentally altering the design premises, namely, the addition of a new module. It seems, however, that the results of the second study suggest that a substantial addition to the design would be required to overcome its presently identified limitations, possibly compromising the fundamentals of a Subsumption design.

What I regard as the fundamental limitation of the design premises for this application appears to me to be related to the notion of reaching an equilibrium in the agent–environment dynamics, a homeostasis. Such homeostasis is, in an important respect, one of the main design goals for a typical embodied Subsumption robot and, pre-dating Brooks, also a goal of cybernetic systems such as Walter's robotic tortoises (Walter 1950) and Ashby's model of the brain (empirically demonstrated with his Homeostat device) (Ashby 1960 [1952]). For these systems, the goal of homeostasis is derived from the idea of biological homeostasis, which is fundamental to the survival of a biological organism (as discussed in section 2.3).

A relation between these systems and free improvisation has been theorised in this thesis on the basis of ecological psychology, which has led to a number of insights. However, the cybernetic premise of dynamic equilibrium or stability does not appear to be as straightforwardly related to the aesthetic-social realm of musical practices. The tendency of the system to return to a narrow range of behaviour, what I conceptualise as its tendency to homeostasis, has been experienced as a drawback in the aesthetic-social realm of free improvisation.

However, with respect to the present study, it is possible to consider some interesting relationships between *following*, pattern recognition, entrainment, and ecological psychology (in particular, the theory of affordances), that point to the ways in which a more complex dynamical system (potentially exceeding the complexity of a typical Subsumption system) may overcome these limitations. Dannenberg (1985) was one of the first to describe systems addressing a computer's responsiveness to human musical performers. Yet a recent framework proposed by Dannenberg et al. (2013) to coordinate studies of human-computer live musical performance primarily conceives of (human and computer) musicians as followers (of tempo, score, soloist, conductor, etc.). This sense of *following* is no doubt a central aspect of many common forms of musical performance. As such, to facilitate more widespread use of such interactive technologies, their project is focused on achieving practical results.

From another perspective, however, it is interesting to consider how such *following* competencies arise from a cognitive standpoint. Large and Kolen (1994, p. 177) view "the perception of metrical structure as a dynamic process where the temporal organization of external musical events synchronises, or entrains, a listeners internal processing mechanisms". Their solution for modelling this phenomenon is to use a network of dynamical systems that can "self-organise temporally structured responses to rhythmic patterns". Doffman (2009) presents an analysis of collective improvisation that supports this view, linking empirical data related to dynamical systems theory with subjective experiential data considered from an ethnographic perspective. The analysis by Doffman suggests that a future version of *Odessa* could coordinate metrical aspects of music with co-performers more effectively than the present version using a cognitive mechanism that would preserve the current architecture (see also Angelis et al. 2013).

Large and Kolen's (1994) view is consistent with that expressed in Port, Cummins, and McAuley (1995), and is similarly consistent with the evolutionary biology perspective of ecological psychology, which views human cognitive function as consisting of biological inheritances from our animal predecessors. Research referred to by Large and Kolen (1994) has shown underlying similarities between visual cortex activity in mammals and complex rhythm perception in humans related to entrainment. Similar to the models of auditory cognition proposed by Port, Cummins, and McAuley (1995) and Large and Kolen (1994), Sloman (2009) offers a sketch of a dynamic model of visual cognition that also links ecological psychology and dynamical systems theory.

Sloman's (2009) criticism of Gibson's notion of affordances is especially interesting in the present context, as it could be applied to an ecological musicology-oriented view of improvisation, in line with that of Clarke (2005a). In response to an identified limitation of Gibson's theory, Sloman (2009, pp. 311–312) introduces the notion of proto-affordances:

> Some animals learn, by playing in the environment, that affordances can be *combined* to form more complex affordances, because *processes can be combined to form more complex processes.* Reasoning about such complex processes and their consequences depends on the ability to combine simpler proto-affordances to form more complex ones. (original emphasis)

With respect to the notion of dramaturgy in collaborative free improvisation discussed in this study, it seems that the real-time guidance of a performance so that it becomes a dramaturgically interesting construction would be facilitated by a performer's conception of emergent musical structures. As structures that emerge over time result from relationships between musical and social processes, the ability to conceptualise these processes as stemming from the past and leading into various potential futures would seem to be crucial to the dramaturgy. As Sloman notes, "the ability to predict future events, and to explain past events [...] is just a special case of a more general ability to combine proto-affordances" (p. 312).

More generally, Sloman (2009, p. 325) states that,

> the ability to perceive not just what is happening at any time but what the possible branching futures are — including, good futures, neutral futures,

> and bad futures from the point of view of the perceivers goals and actions, is an aspect of J.J. Gibson's theory of perception as being primarily about *affordances for the perceiver* rather than acquisition of information about some objective and neutral environment. (original emphasis)

This supports the view of ecological musicology presented in Clarke (2005a). It is also relevant to the discussions presented in earlier chapters concerning the relationship between free improvisation and fixed, notated music, especially with reference to Deliège's theory of cues. Her theory seems to be consistent with Sloman's, as well as with the neurodynamic model of anticipating future musical events presented by Large and Kolen (1994) (for a related neurophysiological account, see Loui et al. 2009; see also Dindo and Chella 2013).

Sloman (2009, p. 325) summarises his critique of Gibson by stating that he does *not* think that Gibson "considered the need to be able to represent, compare and evaluate multi-step branching futures: that would have been incompatible with his adamant denial of any role for representations and computation". Another way to characterise Sloman's theory of proto-affordances is to say that, for Sloman, *imagined possibilities* have potential affordances, which is highly relevant to the present research: if *Odessa* is perceived as an intentional agent, potential interactions with the system are imagined in a difference space of possibilities. That is, if cues for intentional agency perceived in an environment lead to the inference of a present intentional agent, the presence of such an agent in turn affords different possibilities for interaction, including collaboration.

For *Odessa*, the mechanisms producing psychological cues for the perception of intentional agency seem to be effective in establishing it as a valid collaborative partner. And, even with its limited musicality, it also seems to produce the right cues for facilitating the general behaviour of collaborative free improvisation. However, it is clear that the lack of understanding of musical significance, of long-term musical structure, of emotion and dramaturgy, is missing.

While the interplay of layers in *Odessa* allows it to, for example, play material that diverges from a human co-performer, for a typical expert improvisor, such divergence is not merely arbitrary, but is motivated by structural, dramaturgical, or emotional aspects that require a level of understanding both of what other players are doing, and what the whole piece appears to be doing. This is likely not possible without a vastly more complex apparatus — an apparatus typical to human biology and theorised in AI research, but not yet implemented in an artificial system.

The concluding chapter reflects on what has been learned from both studies, from the design and the experiments, and considers possible future research. It also returns to the theoretical themes introduced in previous chapters, and considers how these relate to what has been learned from both studies.

Chapter 7

Conclusions

This research began with a general research question, asking if an artificial agent built using Subsumption could effectively perform human–computer interactive collaborative free improvisation. The assumption that it could was based on several interrelated theoretical perspectives, particularly, the strong link between Subsumption and ecological psychology, and the recent musicological theorisation of improvisation from the perspective of ecological psychology. A key motivation for investigating this question was the prevalence of interactive systems for human–computer collaborative free improvisation, none of which used Subsumption, although Subsumption had been previously shown to have a viable musical application.

On the basis of empirical studies, this general research question was answered affirmatively: *Odessa*, an artificial agent built using Subsumption can effectively support human–computer interactive collaborative free improvisation. In this case, "effectively" is taken to mean that expert improvisors have, overall, experienced a positive collaborative result that demonstrates the competence of the system to engage in collaborative improvisational behaviour. While the system has been shown to have several shortcomings, these do not render the system incompetent; on the contrary, a surprising number of strengths of its parsimonious architecture have been demonstrated, with respect to its ability to collaboratively perform free improvisation with experts.

These strengths are surprising because other comparable systems are significantly more complex than *Odessa*, in their use of extensive cognitive models, machine learning, statistical analysis, musical ontologies, and other stored domain-specific information, for example, pertaining to melody and harmony. *Odessa*, however, as with any Subsumption system, uses no complex internal representations or stored musical knowledge, and does not have a central locus of decision-making; rather, it consists only of a small collection of interconnected low-level simple mechanisms.

From a methodological perspective, I proposed that a Subsumption agent for human–computer collaborative free improvisation could be qualitatively investigated on the basis of the human interaction experience, in particular, to examine the social, psychological, and cognitive processes that underpin collaborative improvisation. I have drawn on qualitative methodologies from similar studies of music and collaborative improvisation, and also described how qualitative study was generally relevant to the evaluation of a Subsumption system. I have demonstrated one way in which an experimental investigation and ethnographically-inspired method could co-exist, as suggested by other investigators of HCI. I have indicated the rigour of the results, with descriptions of the experimental setup, designed to minimise confirmation bias, and with a detailed analysis of participant data. The results obtained from the study have indeed revealed insights into some of the processes that underpin collaborative improvisation, which will be summarised below.

7.1 Revisiting the detailed research questions

Three more specific research questions were posed. The first asked, *Will the system be perceived as an intentional agent by (sonically or musically) exhibiting the appropriate behavioural cues?*

The appropriate behavioural cues were derived from previous empirical psychological research (section 2.5.2). This research suggested that an interplay of cues for agency, as well as cues for accommodation and resistance, would be sufficient to present a behaviour that would be interpreted as that of an intentional agent. The production of these cues was facilitated by the interplay between the layers of *Odessa*'s Subsumption design, namely, the *Play*, *Adapt*, and *Diverge* layers (section 3.2.2.3). The empirical results have shown that *Odessa* was indeed perceived as an intentional agent in the context of collaborative improvisation (Chapter 5).

Further studies would be necessary to determine if these cues served as the exclusive source of such attribution, but the suggestion that they played a significant role is bolstered by the fact that the participants had explicit knowledge that they were interacting with a non-human; moreover, *Odessa* exhibited transparently non-human behaviour. Thus, the attribution of intentional agency could not be due to the mistaking of the system for a human or human-like entity, such as a humanoid animatronic robot, nor could it be due to a convincingly simulated human performance using techniques to disguise its non-human origins.

The second of the more specific research questions asked, *Will the performance be experienced as collaborative if the system demonstrates competence as a free improvisor and is responsive to the human co-performer?*

Contrary to the expectation that the system would always be experienced as collaborative, in some circumstances, the system did not respond adequately or did not demonstrate improvisational competence, despite a general positive outcome in these respects for a majority of the performances (section 5.2.2). When the system did not succeed in these respects, the performance was indeed *not* experienced as collaborative; when it did succeed in these respects, it was typically experienced as collaborative. At this time, it is not clear what further research would be need to better understand how competence is demonstrated in an aesthetic context, outside of merely technical criteria. A similar difficulty applies to investigating how subtle forms of responsiveness are recognised, or what cues might be used in their signalling, in a continuous aesthetic context such as musical improvisation.

The third of these research questions asked, *Will the collaboration be engaging to the human co-performer if the system provides sufficient musical affordances for the human co-performer?*

There was evidence that the affordances did *contribute* to the collaborative engagement, but, ultimately, these affordances were not sufficient to facilitate sustained engagement. Even when such affordances were present, one key shortcoming of the system performance was its inability to demonstrate an understanding of what was taking place in the performance over a longer period of time, on the basis of a sensitivity to dramaturgy, emotion, and other aesthetic-musical qualities (Chapter 6). In my view, this is a constraint of the architecture: it seems that even if it were possible to add to the architecture in such a way that would demonstrate the necessary aesthetic-musical understanding and sensitivity in the system performance, it would, in an important respect, cease to be the same architecture, simply due to the high level of complexity that would be built into the system; it is likely, however, that the classification of the architecture would remain subject to debate.

It would be possible, however, to modify the research question by using an expanded definition of affordance, by saying, for example, that in demonstrating a certain understanding of or sensitivity to the music, the system could afford a more dramaturgical performance (section 6.5). Sloman's (2009) concept of proto-affordances in this context merits further investigation. For the present study, only lower-level, short-term affordances were considered. Thus, from the study on *Odessa*, I would conclude that (short-term) affordances are necessary, but not sufficient for sustained collaborative engagement.

A corollary of these three research questions was the suggestion that the features built into *Odessa* would potentially lead to the perception or attribution of a musical identity or personality that was not specified at design time; rather, it would emerge from the interaction of behavioural layers in the collaborative improvisational agent, and in the overall interaction of the system with a human co-performer. This indeed turned out to be the case, as most participants attributed at least a coherent identity to the system, if not a musical personality (Chapters 5 and 6). Further research could attempt to differentiate between aspects of musical identity or personality that are a result of certain invariances across performances, as opposed to aspects that only manifest in the course of specific collaborative interactions. This could aid in investigations into the relationship between identity/personality and collaboration in aesthetic contexts, as well as in other contexts for collaboration.

7.2 Insights from the study

Given the successes of the first and second iterations of the system, a lot has been learned about the capability of a small collection of simple mechanisms to model the behaviour of a collaborative improvisor. My investigation has considered the role of the contextual framework for interaction, the role of the inferences and interpretations made by collaborating musicians, and some of the cue production mechanisms that facilitate these inferences. I have shown that a parsimonious Subsumption system can achieve a complex and robust musical interaction that comes surprisingly close to human-level expertise without the use of elaborate, sophisticated, and expensive computation. In particular, *Odessa* achieves its performance without the use of machine learning, probabilistic analysis, or formal musical knowledge.

Research on *Odessa* supports the idea that in-the-moment inferences, based on behavioural cues perceived in real time, can lead to the attribution of intentional agency. Furthermore, the fact that the musical behaviour exhibited by *Odessa* was typically regarded as musically coherent supports another aspect of perceptual cue theory: the notion that musical cues can lead to inferences regarding musical structures and relationships that are not necessarily formally encoded or deliberately enacted in the formulation or production of material. Cues are effective relative to an interpretive context, which is in line with the ecological view that agents respond to different aspects of their environment depending on what is relevant to them at a given moment. Subsumption robots, for example, in their agent–environment interaction, display an ecological sense of “intelligence” by responding to the environmental cues that are relevant to their performance. Significantly, *Odessa* provides a basis for extending an ecological theory of cues to an environment containing other agents, that is, for agent–agent interaction. Thus, this research interestingly ties together superficially unrelated research in human developmental psychology, cognitive ethology, and music perception theory, and, more generally, also connects to topics in robotics, AI, cognitive science, and neuroscience.

To summarise the general claims about human behaviour in collaborative free improvisation that are suggested by the present research:

- environmental cues may be perceived in such a way as to suggest the presence of an intentional agent;
- a perceived intentional agent provides an opportunity for collaborative interaction;
- the agent may provide affordances that present multiple possibilities for musical response by a co-performer;
- the agent's responses to the co-performer may demonstrate its competence (or lack thereof). Some of the possible responses include the agent providing stimulating challenges to a co-performer, serving in a role that appears to be musically supportive, or providing a parallel independent part that appears to function together musically with that of a co-performer;
- finally, because the agent and co-performer are continuously mutually responsive in the interaction, responses must (at least at times) suggest a recognition of the accumulated history of responses. This mutually responsive interaction (with a history) can be described as a collaborative exploration or development of musical material (e.g., remaining with something just played before a significant change is introduced, avoiding something previously played, returning to something previously played, etc.).

This research has also revealed unexpected insights into the ethnomusicology of free improvisation. The conception that free improvisation can be defined simply as the absence of certain musical features, most notably, style, clearly does not provide an adequate account of the practice. First off, it can be noted that positively defined features of free improvisation can be observed, namely, those that relate to the skilful negotiation of traditionally musical phenomena like melodies and harmonies, or to other musical facets such as texture. But, even if one were to say that this skilful negotiation itself takes place in the absence of style, which is its defining characteristic, one risks the criticism that a historically established approach to this negotiation has, by now, crystallised into a style.

This criticism, however, is short-sighted, because it regards the negotiation as a narrowly defined activity. One must account not only for the negotiation of outwardly musical material, but also for the collaborative negotiation of nominally extramusical aspects of the performance in its social, emotional, and dramaturgical dimensions. Freely improvised music may, but need not, follow, for example, patterns of tension and release; while such patterns are indeed associated with particular musical traditions, they seem to be at a different level of abstraction than what is traditionally regarded as style. The fact that such improvisation is "free" not only relates to an ideal freedom from musical style, but also to a practical freedom from the requirement to recognisably fit any previously established form of musical structure (while, of course, remaining free to do so).

The contributions of this research can be summarised as follows:

- I applied Subsumption in a novel way to a real-time human-interactive musical agent (a novel application with respect to Subsumption systems generally and to interactive improvisation systems);
- I demonstrated that cognitive architecture plays a greater role in free improvisation than merely facilitating the adherence to explicit or implicit cultural rules, in particular, by the ways in which it integrates perceptual cue information that connects to and affects action;
- I contributed to the empirical grounding of Clarke's (2005a) ecological theory of music perception and meaning-making; and,
- I developed a novel methodological approach to investigating aspects of music psychology and music cognition using computational modelling (interactive AI) with human participants.

7.3 Future work

For the next stage of this research, I plan to develop a biomimetic version of the current behaviour-based model. This next version will make use of longer-term memory for perceptual cue abstraction, likely using neural networks for pattern recognition. Furthermore, some notion of proto-affordances should also be accounted for by the model, giving it the ability to represent multiple possible pathways within the musical environment, that is, to flexibly represent different connections between past and future. These pathway representations would be ultimately separated into perceptual and cognitive stages. A selection mechanism will also be required, to link particular pathways with action potentials in what could be considered the intentional arc of the system.

7.4 Beyond *Odessa*

Odessa has served as an interesting means for investigating collaborative improvisational experience. But what would it mean for *Odessa* to be a truly collaborative improvisor, in the way a human can be? While one can model aspects of personal competence, or the competencies of others, this, however, also means that system ultimately depends upon following someone else's model, rather than developing its own capacities. Given current technology, it is difficult to imagine an alternative; certainly those working within these constraints have developed aesthetically-rewarding systems — such as some of those discussed here — that are themselves works of art, and that continue to enrich our culture.

From a speculative perspective, however, on the basis of the research on *Odessa*, I believe there is a way of developing a collaborative improvisor that more closely approximates the way an autonomous human improvisor operates. In particular, there are four points I would like to raise concerning what is missing from *Odessa*; thus, the focus is on free improvisation, but the points may also apply more generally. Three of the four components missing from *Odessa* are *in principle* technologically achievable at present, although I do not believe any successful implementation of all of them (in this context) is imminent. A fourth missing component, however, although also *in principle* possible in a technological implementation, is, in my opinion, unlikely to be seen for a considerable amount of time.

The first three, simply put, are perception, knowledge, and the ability to recognise significance without its prior specification. Specific to this context, I believe that, regarding the first point, a perceptual apparatus for such an agent must include an on-going sensitivity to spatially-located sounds, such as those of itself and other performers, as well as to the collective sound, and how all of these relate to the acoustic space with which they are constantly interacting. Current technology allows for some of these features, but no current technology that I am aware of is capable of taking advantage, as a performer would, of such a sensory apparatus in an arbitrary, complex, real-world environment with multiple sound sources, i.e., a human-level hearing apparatus. (This relates to research on computational auditory scene analysis. For a relatively recent account of the state of the art, see Wang and Brown 2006; for some of the difficulties in achieving a human-level system, see Bregman 2013.)

The second component, knowledge, is also clearly a feature of many or even most current AI systems. However, no system I am aware of is capable of performing improvised music in real time while, in addition, *on the basis of knowledge* of multiple human cultures and traditions of music, either typically avoid musically invoking any of these cultures (e.g., their melodic, harmonic, or rhythmic patterns); or, deliberately make subtle or implied references to them. Many expert free improvisors are trained in one or more traditions, whereas from an engineering perspective, it would seem more effective to simply encode specific patterns to prevent them from being fully reproduced. One could imagine, however, an artificial neural network trained on a corpus containing multiple musical traditions, designed to produce musical output without coinciding with certain features of its training sets, which, to my knowledge, has not yet been tried. Whether the resulting system would approximate human competence is a matter for further discussion.

The third component, the ability to recognise significance without its prior specification, is problematic for AI systems. An artificial improvisor must be able to recognise the significance of what other performers are doing, even if the significant features to be detected have not been specified to the system. To give an example, if a performer finds certain significance in their own 'part', and another performer responds to this part according to some other significant features that were not recognised by the first performer, the first performer may in turn come to find new significance in their own part based on the way in which the second is responding. This is possible even if the first performer has no advanced knowledge of the significant aspects that may be identified by the second performer.

The fourth component has not yet been named: it is the development of an aesthetic sensibility and faculty of judgement. In my opinion, this is not within reach, and many advances must be made before coming close. Importantly, this sensibility and faculty of judgement are not adequately understood when viewed as simply imposed from the outside; for humans, both are the result of life experience, reflection, and the related personal evolution of one's decision-making capacities. According to ethnographic accounts of free improvisors, the questions of when to be silent, when to support which player or piece of musical structure, when to provide a complementary or antagonistic opposition, and so on, generally do not appear to be determined randomly, not by an algorithm or formula, not by a conscious statistical analysis, and not by a sampling of listener preferences, but by a judgement made within a perceived situation. For humans, a developed faculty of judgement is not the blind application of a maxim, but a capacity that develops over time, as it is influenced by both domain-specific experience, for example, in performing music, but also by domain-general experience, such as the encounters and decisions of everyday life; these also shape one's aesthetic sensibilities. In my opinion, until a machine can experience a related sense of everyday life, and undergo a similar process of reflection and development, a machine improvisor cannot be said to improvise in a manner equivalent to humans (see Linson 2013).

7.5 Concluding remarks

This study has considered cognitive aspects of collaborative improvisational behaviour from an ecological perspective, as they relate to psychological and social processes. From a broader perspective, the study also complements other research into cognition. For example, Bengtsson et al. (2007) believe that their neuroscientific study on the cortical regions involved in piano improvisation "demonstrates that musical improvisation may be a useful behaviour for studies of the neurocognitive processes underlying an ecologically relevant creative behaviour". The authors note that "an important next step will be to analyse the neural underpinnings of the cognitive components of improvisation". Thus, we must begin with an empirically grounded understanding of such cognitive components, to which the research on *Odessa* contributes. While the intertwining of mind, brain, body, and environment continues to be theoretically explored, we are increasingly able to comprehend the extensiveness of their mutual dependencies. Perhaps this will lead us to recognise the interdependencies within our social and natural environments in a way that will facilitate greater participation in the collaborative improvisation of our shared future.

Appendix A

On the implementation

A.1 Using ChucK for *Odessa*

ChucK (version 1.2.1.3 for GNU/Linux), a programming language dedicated to music and other audio applications, was selected as the language in which to implement *Odessa* based on several key strengths. I introduce these strengths here and present further details about each below. In particular, ChucK:

- runs in a portable virtual machine designed to support concurrent processes;
- is designed to support rapid prototyping; and,
- features native libraries for audio processing.

After explaining the significance of these strengths to developing *Odessa*, I also explain some programming decisions made with ChucK in relation to handling audio. This is followed by a description of using ChucK for Subsumption (section A.2).

A.1.1 Virtual machine and concurrency

ChucK runs in its own virtual machine, which allows code to be run on different physical machines without extensive configuration. As long as the virtual machine compiles successfully and the same version is used, compatibility issues do not arise between physical machines. This was a significant factor in selecting the software for a multi-year research undertaking, for which it is impossible to guarantee the continued use of a single functional physical machine.

In addition, for envisioning a Subsumption implementation, the ChucK virtual machine offers a crucial advantage to other programming environments. Namely, it facilitates the ability to run many processes concurrently which "behave as if they are running in parallel" (Wang 2008, p. 58). Although this is in principle common in other programming languages, as Fiebrink et al. (2008, p. 155) point out, ChucK's "sample-synchronous concurrency would be extremely challenging and/or inefficient in C++, Java, or a library built in these languages, due to their support for pre-emptive, thread-based concurrency, and the consequent need to contend with thread instantiation, synchronization, bookkeeping, etc.". (This is discussed again below in the context of rapid prototyping.) The way ChucK handles processes is useful for developing a Subsumption system, for organising software modules into concurrently running networks of modules. Another feature of Subsumption, namely, that certain parts of the system typically operate independently without sharing data (Brooks 1999, p. 26), is also possible in ChucK. Processes in ChucK, called 'shreds', are such that "any given shred does not necessarily need to know about other shreds" (Wang 2008, p. 58).

A.1.2 Rapid prototyping

The author and chief architect of ChucK, Ge Wang, writes that,

> audio programming, in both computational acoustics research and in music composition and performance, is necessarily an experimental and empirical process; it requires rapid experimentation, verification/rejection/workshopping of ideas and approaches, and takes the forms of both short-term and sustained prototyping. (Wang 2008, p. 3)

As Wang (2008) states specifically of developing ChucK, "we believe that rapid prototyping, in and of itself, is a uniquely useful approach to programming audio, with its own benefits (and different ways of thinking about problems)" (p. 3). Elsewhere, he refers to "the rapid prototyping mentality where continuous exploration and experimentation are valued" (p. 115). Although a programming language may be used for purposes unforeseen by the designer of the language, it is nevertheless the case that the language designer's perspective on how the language should be used is likely to be reflected in certain features of the language.

Brooks speaks similarly of the mentality behind formalising the Subsumption approach, and how the architecture was expected to affect future robot development. A Brooksian robot development process begins with early deployment in an unsimplified target environment, so that real challenges to system performance can be taken into account.

An iterative process of extensive testing, adjusting, debugging, and fine-tuning of each system behaviour is undertaken until further development commences (Brooks 1999, pp. 91–92). This is opposed to a traditional development approach in which a long-term, complex construction is completed prior to initial deployment. The traditional approach often results in uncertainty about the potential sources of system errors, and thus a similarly long time may be needed to find and fix defects in each successive iteration.

The fact that Subsumption calls for short development cycles was a consideration in selecting a language tailored for rapid prototyping. Beyond these successive cycles, however, it was also necessary to consider the overall time constraints on this research. When prototyping *Odessa* under these constraints, several advantages anticipated by the lead developers of ChucK proved accurate. In particular, Fiebrink et al. (2008, p. 155) summarise what they regard as the three main advantages of using ChucK with limited development time:

1. *ChucK is effective for short-term audio software development:*

   > ChucK implicitly manages real-time audio and buffering, and the language is tailored for audio, so [...] system implementation time and code length are greatly reduced.

2. *ChucK lends itself to "rapid turnaround experimentation":*

   > Through the application of on-the-fly programming and ChucK's concise audio programming syntax, one can quickly prototype systems and subsystems, changing parameters as well as the underlying system structure, and experience the results almost immediately.

3. *ChucK's audio-centred concurrency model facilitates shorter development times:*

   > Parallel management of [audio] feature extraction, even at multiple rates, is straightforward in ChucK. This sample-synchronous concurrency would be extremely challenging and/or inefficient in C++, Java, or a library built in these languages. (Fiebrink et al. 2008, p. 155)

These advantages proved useful to developing the initial *Odessa* system, as well as to the short development cycles called for by Subsumption in the process of progressively enhancing the system's capabilities.

A.1.3 Native audio processing libraries

For *Odessa*'s real-time parsing of the audio stream from the musical instrument played by the human participant, it was possible to use native ChucK objects and libraries. As Wang et al. (2007) state, ChucK "supports straightforward extraction of arbitrary time and frequency domain features, which might be stored or used as parameters to drive real-time synthesis" (p. 39). ChucK has native support for extracting several standard spectral features, and can extract features at separate rates (p. 39). For *Odessa*, the frequency in Hertz of the input signal is extracted in real-time to approximate pitch, and the RMS (root-mean-square) amplitude[1] is also extracted in real-time to approximate loudness.[2] ChucK also provides convenient native functions for generating pseudorandom numbers and conversion between common value types (e.g., between RMS amplitude and decibels).[3]

A.1.4 Synthesis and MIDI

In Chapter 4, the motivations behind the decision to use piano sounds for *Odessa* are discussed. In this section, a more technical account of *Odessa*'s sound production is given. ChucK conveniently provides native support for both audio synthesis and the Musical Instrument Digital Interface (MIDI) protocol. For *Odessa*, the MIDI protocol provides a way to interface with specialised hardware such as an electromechanically controlled acoustic piano (e.g., Yamaha Disklavier). While ChucK's synthesis capability can be used to produce continuous as well as discrete pitches, the decision to use piano sounds for *Odessa* meant continuous pitches were not necessary. For this reason, rather than the native ChucK synthesis, an unofficial extension for ChucK was used to add support for a special file format for synthesisers called *SoundFonts*, that enable the use of instrument samples as a sound source. This allowed for piano samples to be used for *Odessa* when a Disklavier was not available, without requiring elaborate synthesis.[4]

For *Odessa*, frequency values extracted from input (discussed in section 3.2.3) have been converted to the nearest discrete piano pitches, partly using a native ChucK

---

[1] "The root mean square value [...] is representative of the energy in the signal and is computed by squaring the signal, averaging over some time period and taking the square root" (Cabot 1999, p. 742).

[2] Loudness is technically dependent on the physical position of the listener with respect to the sound source, as well as on the listener's "psychological and physiological conditions, with reference to fatigue, attention, alertness, etc.", not merely on the amplitude or corresponding intensity of the signal (Fletcher and Munson 1933). See also Fletcher (1934).

[3] For further details, see the ChucK documentation at `http://chuck.cs.princeton.edu/doc/`.

[4] On the ChucK mailing list *chuck-users*, Ge Wang notes that, for ChucK, "there is a prototype fluidsynth UGen [sound-producing unit] in development (by Kyle Super) that loads and plays soundfonts" (`https://lists.cs.princeton.edu/pipermail/chuck-users/2008-April/002759.html`). In a later message to the list, Tom Lieber includes Kyle Super's experimental patch (`https://lists.cs.princeton.edu/pipermail/chuck-users/2008-October/thread.html#3384`). At present, neither this patch nor other SoundFont support has been included in the official ChucK release.

frequency-to-MIDI value conversion function, and partly using original supplementary code to handle pitches above and below the piano keyboard. The supplementary code recursively raises or lowers values by an octave until they fit into the piano pitch range. Since the current version of *Odessa* is configured to conform to MIDI output (for use with a Disklavier or similar device), the approximate loudness measurement using RMS amplitude is converted to a 7-bit MIDI velocity value (the MIDI standard for volume, referring to the velocity with which a keyboard note is struck). Thus, in further system descriptions below, when referring to loudness, if it is specifically relevant, the term *amplitude* is used for input, and *velocity* for output.

Before any performance with *Odessa*, a sound check is required to calibrate the system. During the sound check, I establish a maximum and minimum RMS amplitude of the input signal (provided by the human performer), and map this range to 7-bit values, where the maximum is 127 and the minimum is 50. Values 0–49 are treated as silence, that is, as if there is no audible input. In practice, this means that barely audible sounds (rather than strictly inaudible ones) may fall below the threshold, but this has proved to be an acceptable tolerance in this research. (These ranges were established in relation to *Odessa*'s output capacity, discussed further below. With either synthesised output or a Disklavier, it was empirically determined that values below 50 were typically inaudible to the co-performer; see Risset and Van Duyne 1996 for more on the "idiosyncrasies of the Disklavier", especially pertaining to loudness and velocity.)

A.2 Coding Subsumption in ChucK

A.2.1 Modules and networks

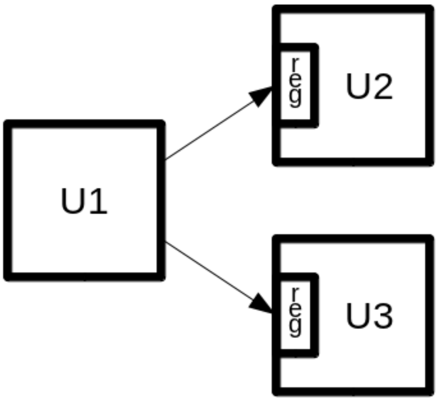
As described above, Subsumption robots are built by combining independent software modules into networks. These networks operate as parallel layers that each implement a simple behaviour. The layers are also interconnected to implement more complex behaviours, or such complex behaviours may emerge from concurrently operating parallel layers without explicit interconnections. ChucK allows for individual files to be loaded into its virtual machine (VM), which are compiled at run-time and executed as processes called *shreds*. For *Odessa*, the code for each Subsumption module is contained within its own ChucK file. All the files are loaded into the VM in a single batch and simultaneously executed as independent shreds.

One initial problem with using ChucK for Subsumption is that, in the ChucK version (v1.2.1.3) used for *Odessa*, it is not possible to directly pass a message from one shred (module) to another. However, a workaround makes possible the initialisation of global variables (global to all shreds in the VM), such that all concurrent shreds have read and write access to them. It would, in principle, be possible to use these global variables for a blackboard architecture (Figure 7c), which Brooks (1999, p. 97) notes is conceptually different from Subsumption. As he describes, a blackboard system tends to "make heavy use of the general sharing and availability of almost all knowledge". Moreover, a typical blackboard abstraction hides "from a consumer of knowledge who the particular producer was".

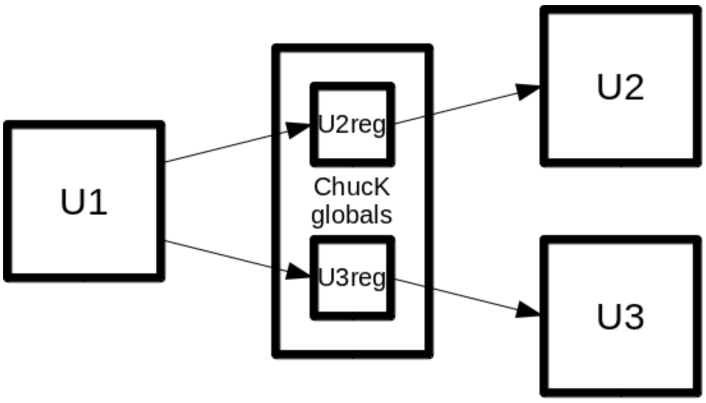
For *Odessa*, the Subsumption strategy was followed rather than using a blackboard, such that the global variables are organised as virtual local registers to each module (Figure 7b). To clarify, in a physical Subsumption robot, there may be a physical module (call it U1) that is connected by wires to two other physical modules (U2 and U3). U1 may send a message over one of the wires that has as its destination a register of U2 or U3 (Figure 7a). When a register is written, its existing contents are replaced, and the arrival of a message can trigger a state transition. For example, U1 may send a message to a register of U2, causing a state change in U2, but leaving U3 unaffected.

In coding *Odessa*, the global variables, although technically stored in their own globally accessible module, are assigned a unique destination in a particular target module. For example, using the same naming convention from the above example, a global variable called *U2reg*, while not internal to the U2 module (as this is not possible in ChucK), only receives messages for U2; they are read only by U2, and trigger a state transition only in U2 (Figure 7b). As a means of further preserving the Subsumption abstraction, native ChucK event-based triggers are used. An event-based trigger is such that as soon as a value is written to an assigned global variable (e.g., U1 writes to *U2reg*), the intended recipient (e.g., U2) is signalled to read the (*U2reg*) value. In this way, it is as if the message has been passed directly from (e.g.) U1 to U2, even though the line of transmission in ChucK is technically indirect.

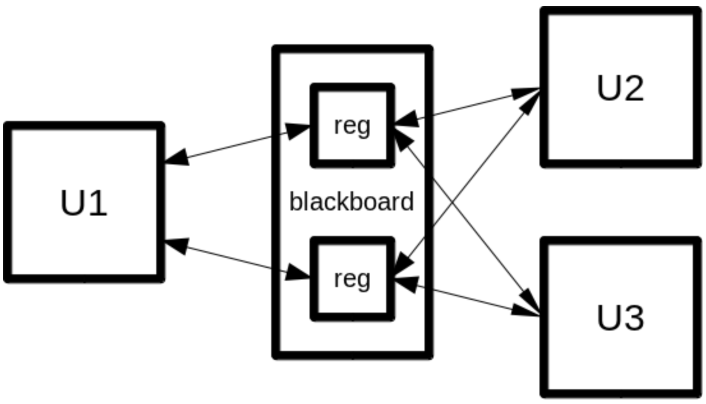
Brooks (1999) refers to message passing between modules as being confined to a low number of bits. He specifies that messages typically have 8- or 16-bit values, although in some instances values range from as low as 1-bit to as high as 32-bit, depending on their purpose (for physical robots, this size also determines what hardware is needed) (Brooks 1999, p. 172). Thus, in principle, a future version of *Odessa* could use 16-bit messages to pass high-resolution information about frequency in Hertz and fine-grained amplitude measurements between modules, without compromising the basics of the Subsumption approach. The current implementation of *Odessa* is even more conservative. Frequency and amplitude are converted to 7-bit MIDI values before being passed as messages, such that only a maximum 7-bit message size is used.



(a) Subsumption message passing with hardware. Connections are explicit and permanent; registers are physically in modules.



(b) Subsumption message passing with *Odessa*/ChucK. Connections are explicit and permanent; ChucK limitation forces workaround using global variables, but Subsumption abstraction is maintained.



(c) Message passing with a blackboard architecture. Data is shared by all modules.

Figure 7: Message passing comparison (arrows indicate data flow).

As stated above, for each Subsumption module, the arrival of a message or the expiration of a timer can trigger a state transition (Brooks 1999, p. 41). These timers (also called 'alarm clocks' in Subsumption) are straightforward to implement in ChucK, which has a native ability to measure temporal duration in units such as milliseconds. Thus, a state transition in a module can be triggered by a timer, that is, when a specified amount of time has passed.

A.2.2 Synchronous versus asynchronous clock rates

The ability of the modules of a system to function together asynchronously is another characteristic of Subsumption. However, this is not possible with ChucK. While ChucK can receive asynchronous input from an external device, internal processes in a ChucK VM run synchronously by design. Nevertheless, to adhere to Subsumption design principles, *Odessa* is programmed such that it is not *dependent* on any synchronisation between modules to achieve its performance. The remainder of this subsection will explain this premise.

Generally speaking, ChucK processes running in the VM are synchronised to a global clock rate. The same is technically true of all processes running in an operating system (OS), which has a global clock rate determined by its hardware, although, from a user perspective, this low-level synchronisation may not appear to be significant. Thus, in a typical scenario, any software running on a single machine is ultimately synchronised to a single clock. In contrast, one would expect a robot with physically separate microprocessors linked together by wires to run concurrent processes with asynchronous clock rates, unless the microprocessors were slaved to a master clock.

As Subsumption robots are not dependent on modules running synchronously, they do not require a master clock. As Brooks (1999, p. 173) states, "there is no explicit synchronisation between a producer and a consumer of messages". In part, the asynchronous performance of a Subsumption system is achieved by having modules routinely discard incoming messages (e.g., when they arrive at a high rate), and by using internal timers to initiate state transitions (e.g., when messages arrive at a slow rate or do not arrive at all). Thus, Subsumption modules are designed to effectively cope with missed data; that is, their operation is not dependent on the arrival of every data packet. In fact, Brooks (1999, p. 173) notes that it is common in some cases for only one tenth of messages to be examined by a receiver, since "message reception buffers can be overwritten by new messages before the consumer has looked at the old one". In addition, a module may not be dependent on the arrival of *any* packet, as it may initiate actions independently without incoming data, triggered by the expiry of timers.

Odessa is built according to this same data flow logic, such that the synchronous behaviour of the ChucK VM can be regarded as irrelevant to the system performance. In future research, to demonstrate its asynchronous capabilities, *Odessa* could be rebuilt using a separate physical machine for each module; modules would be physically wired together for message passing, as with a physical Subsumption robot. For the present research, however, the practical benefits of running *Odessa* on a single machine outweigh the potential gain of establishing experimentally that it would remain viable if distributed across multiple machines.

Appendix B

Audio examples

The following supplementary audio can be found at
`http://percent-s.com/linson-phd-thesis-audio`

1. *Odessa* (1st prototype, Disklavier):
   Duet w/ Adam Linson (double bass), 4m43s.
   Live at Interactive Keyboard Symposium, Goldsmiths, University of London, 10 November, 2012 (referenced in section 4.1.1.3).

2. *Odessa* (2nd prototype, Disklavier):
   Duet w/ Evan Parker (soprano saxophone), 8m59s.
   Follow-up study (Chapter 6), 21 August, 2013.

3. *Odessa* (2nd prototype, Disklavier):
   Duet w/ John Russell (guitar), 15m20s.
   Follow-up study (Chapter 6), 21 August, 2013.

4. *Odessa* (2nd prototype, Disklavier):
   Trio w/ Evan Parker (soprano saxophone) + John Russell (guitar), 9m11s.
   Follow-up study (Chapter 6, see especially section 6.4.6), 21 August, 2013.

Bibliography

Agre, P. E. Computational research on interaction and agency. *Artificial intelligence*, 72(1):1–52, 1995.

Agre, P. E. and Horswill, I. Cultural support for improvisation. In *AAAI-92 Proceedings*, pp. 363–368, 1992.

Agre, P. E. and Rosenschein, S. J., editors. *Computational Theories in Interaction and Agency*. Cambridge: MIT Press, 1996.

Angelis, V., Holland, S., Upton, P. J., and Clayton, M. Testing a computational model of rhythm perception using polyrhythmic stimuli. *Journal of New Music Research*, 42 (1):47–60, 2013.

Ashby, W. R. Requisite variety and its implications for the control of complex systems. *Cybernetica*, 1(2):83–99, 1958.

Ashby, W. R. *Design for a brain: The origin of adaptive behaviour*. London: Chapman & Hall, 1960 [1952].

Assayag, G., Bloch, G., Chemillier, M., Cont, A., and Dubnov, S. Omax brothers: A dynamic topology of agents for improvization learning. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pp. 125–132, 2006.

Babbitt, M. Twelve-tone rhythmic structure and the electronic medium. *Perspectives of New Music*, 1(1):49–79, 1962.

Bailey, D. *Improvisation: its nature and practice in music*. Cambridge: Da Capo Press, 1993 [1980].

Baldwin, D. and Baird, J. Discerning intentions in dynamic human action. *Trends in cognitive sciences*, 5(4):171–178, 2001.

Balhorn, M. The rise of epicene they. *Journal of English Linguistics*, 32(2):79–104, 2004.

Ball, L. and Ormerod, T. Putting ethnography to work: the case for a cognitive ethnography of design. *International Journal of Human-Computer Studies*, 53(1): 147–168, 2000.

Barber, M. N. and Ninham, B. W. *Random and restricted walks: Theory and applications*, volume 10. London: Routledge, 1970.

Barrett, J. L. and Johnson, A. H. The role of control in attributing intentional agency to inanimate objects. *Journal of Cognition and Culture*, 3(3):208–217, 2003.

Beer, R. D. Toward the evolution of dynamical neural networks for minimally cognitive behavior. In Maes, P., editor, *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pp. 421–429, Cambridge, 1996. MIT Press.

Bengtsson, S. L., Csíkszentmihályi, M., and Ullén, F. Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of cognitive neuroscience*, 19(5):830–842, 2007.

Beyls, P. Interaction and self-organisation in a society of musical agents. In *Proceedings of ECAL 2007 Workshop on Music and Artificial Life (MusicAL 2007)*, 2007. URL `http://cmr.soc.plymouth.ac.uk/Musical2007/papers/Beyls.pdf`. Accessed: 2 Aug 2013.

Biles, J. Genjam: A genetic algorithm for generating jazz solos. In *Proceedings of the International Computer Music Conference*, pp. 131–137, 1994.

Blackwell, T. Swarm music. Master's thesis, University College London, 2001.

Blackwell, T. Swarming and music. In *Evolutionary Computer Music*, pp. 194–217: Springer, 2007.

Blackwell, T. and Jefferies, J. Swarm tech-tiles. In *Applications of Evolutionary Computing*, pp. 468–477: Springer, 2005.

Blackwell, T. and Young, M. Self-organised music. *Organised Sound*, 9(2):123–136, 2004a.

Blackwell, T. and Young, M. Swarm Granulator. In Raidl, G. et al., editors, *Applications of Evolutionary Computing: EvoWorkshops 2004, LNCS*, volume 3005, pp. 399–408. Berlin: Springer, 2004b.

Blackwell, T. M. and Bentley, P. Improvised music with swarms. In *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, volume 2, pp. 1462–1467. IEEE, 2002a.

Blackwell, T. M. and Bentley, P. Don't push me! collision-avoiding swarms. In *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, volume 2, pp. 1691–1696. IEEE, 2002b.

Borgo, D. and Goguen, J. Rivers of consciousness: The nonlinear dynamics of free jazz. In Fisher, L., editor, *Jazz Research Proceedings Yearbook*, pp. 46–58, 2005.

Braasch, J. A cybernetic model approach for free jazz improvisations. *Kybernetes*, 40 (7/8):984–994, 2011.

Braasch, J., Van Nort, D., Oliveros, P., Bringsjord, S., Govindarajulu, N. S., Kuebler, C., and Parks, A. A creative artificially-intuitive and reasoning agent in the context of live music improvisation. In *Proceedings of Music, Mind, and Invention: Creativity at the Intersection of Music and Computation*, 2012.

Braitenberg, V. *Vehicles: Experiments in synthetic psychology.* Cambridge: MIT press, 1984.

Bregman, A. S. Three directions in research on auditory scene analysis. In *Proceedings of Meetings on Acoustics, Vol. 19, 010021*, ICA 2013, Montreal, 2013.

Brooks, R. A. New approaches to robotics. *Science*, 253(5025):1227–1232, 1991.

Brooks, R. A. *Cambrian intelligence: The early history of the new AI.* Cambridge: MIT Press, 1999.

Brooks, R. A. *Robot: The future of flesh and machines.* London: Allen Lane, 2002.

Brown, J. In the zone: an autoethnographic study exploring the links between flow and mindfulness for a piano accompanist. *Studies in Learning, Evaluation, Innovation and Development*, 8(2):83–95, 2011.

Bryson, J. The Reactive Accompanist. Master's thesis, Department of Artificial Intelligence, University of Edinburgh, 1995.

Byrne, C., MacDonald, R., and Carlton, L. Assessing creativity in musical compositions: Flow as an assessment tool. *British Journal of Music Education*, 20(3):277–290, 2003.

Castelli, F., Frith, C., Happé, F., and Frith, U. Autism, asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125(8): 1839–1849, 2002.

Chamberlain, A., Crabtree, A., Rodden, T., Jones, M., and Rogers, Y. Research in the wild: understanding 'in the wild' approaches to design and development. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, Newcastle Upon Tyne, United Kingdom, pp. 795–796, New York, NY, USA, 2012.

Chiel, H. J. and Beer, R. D. The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in neurosciences*, 20 (12):553–557, 1997.

Clark, A. *Being There: Putting Brain, Body, and World Together Again.* Cambridge: MIT Press, 1997.

Clark, A. and Chalmers, D. The extended mind. *analysis*, 58(1):7–19, 1998.

Clarke, E. Improvisation, cognition and education. In Paynter, J., Howell, T., Orton, R., and Seymour, P., editors, *Companion to Contemporary Musical Thought. Volume 2*, pp. 787–802. London: Routledge, 1992.

Clarke, E. Empirical methods in the study of performance. In Clarke, E. and Cook, N., editors, *Empirical musicology: Aims, methods, prospects*, pp. 77–102. Oxford: Oxford University Press, 2004.

Clarke, E. *Ways Of listening: An ecological approach to the perception of musical meaning.* Oxford: Oxford University Press, 2005a.

Clarke, E. Creativity in performance. *Musicae Scientiae*, 9(1):157–182, 2005b.

Clarke, E. and Cook, N., editors. *Empirical musicology: Aims, methods, prospects.* Oxford: Oxford University Press, 2004.

Clayton, M., Sager, R., and Will, U. In time with the music: The concept of entrainment and its significance for ethnomusicology. In *European meetings in ethnomusicology*, volume 11, pp. 3–142, 2005.

Collins, H. M. *Artificial experts: Social knowledge and intelligent machines.* Cambridge: MIT Press, 1990.

Collins, N. *Towards Autonomous Agents for Live Computer Music: Real-time Machine Listening and Interactive Music Systems.* PhD thesis, University of Cambridge, 2006.

Conklin, D. and Witten, I. H. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.

Connell, J. A colony architecture for an artificial creature. Technical report, DTIC Document, 1989.

Cope, D. *Computer Models of Musical Creativity.* Cambridge: MIT Press, 2005.

Corkill, D. D. Blackboard systems. *AI expert*, 6(9):40–47, 1991.

Csibra, G. Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2):705–717, 2008.

Csíkszentmihályi, M. *Creativity: The Psychology of Discovery and Invention.* New York: HarperCollins, 2013 [1996].

Dannenberg, R. B. An on-line algorithm for real-time accompaniment. In *Proceedings of the 1984 International Computer Music Conference, Paris*, pp. 193–198, 1985.

Dannenberg, R. B., Jin, Z., Gold, N. E., Sandu, O.-E., Palliyaguru, P. N., Robertson, A., and Stark, A. Human-computer music performance: From synchronized accompaniment to musical partner. In *Proceedings of Sound and Music Computing Conference 2013, Stockholm, Sweden*, pp. 136–141, 2013.

Davidson, J. W. Music as social behavior. In Clarke, E. and Cook, N., editors, *Empirical musicology: Aims, methods, prospects*, pp. 57–75. Oxford: Oxford University Press, 2004.

Davidson, J. W. and Good, J. M. Social and musical co-ordination between members of a string quartet: An exploratory study. *Psychology of Music*, 30(2):186–201, 2002.

Dean, R. T. *Hyperimprovisation: Computer-interactive sound improvisation*: AR Editions, Inc., 2003.

Deliège, I. Introduction: Similarity perception – categorization – cue abstraction. *Music Perception: An Interdisciplinary Journal*, 18(3):233–243, 2001.

Deliège, I. Similarity relations in listening to music: How do they come into play? *Musicae Scientiae*, 11(Discussion Forum 4A):9–37, 2007.

Deliège, I., Mélen, M., Stammers, D., and Cross, I. Musical schemata in real-time listening to a piece of music. *Music Perception*, 14(2):117–159, 1996.

Dennett, D. *The Intentional Stance.* Cambridge: MIT Press, 1987.

Dindo, H. and Chella, A. What will you do next? A cognitive model for understanding others intentions based on shared representations. In Shumaker, R., editor, *Virtual Augmented and Mixed Reality: Designing and Developing Augmented and Virtual Environments, LNCS 8021*, pp. 253–266. Berlin: Springer, 2013.

Doffman, M. Making it groove! Entrainment, participation and discrepancy in the 'conversation' of a jazz trio. *Language & History*, 52(1):130–147, 2009.

Doffman, M. Jammin' an ending: Creativity, knowledge, and conduct among jazz musicians. *twentieth-century music*, 8(2):203–225, 2011.

Dourish, P. *Where The Action Is: The Foundations Of Embodied Interaction.* Cambridge: MIT Press, 2001.

Dourish, P. Implications for design. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 541–550, 2006.

Dreyfus, H. L. *What computers still can't do: A critique of artificial reason.* Cambridge: MIT Press, 1992 [1972].

Dreyfus, H. L. Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. In Husbands, P., Holland, O., and Wheeler, M., editors, *The mechanical mind in history*, pp. 331–371. Cambridge: MIT Press, 2008.

Dreyfus, H. L. and Dreyfus, S. E. Peripheral vision expertise in real world contexts. *Organization studies*, 26(5):779–792, 2005.

Eisenstein, S. *The Film Sense.* New York: Harcourt, Brace, & Co., 1947.

Ensmenger, N. Is chess the drosophila of artificial intelligence? A social history of an algorithm. *Social Studies of Science*, 42(1):5–30, 2011.

Ernst, M. and Bülthoff, H. Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169, 2004.

Fiebrink, R., Wang, G., and Cook, P. Support for MIR prototyping and real-time applications in the ChucK programming language. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pp. 153–60, 2008.

Fletcher, H. Loudness, pitch and the timbre of musical tones and their relation to the intensity, the frequency and the overtone structure. *The Journal of the Acoustical Society of America*, 6(2):59–69, 1934.

Fletcher, H. and Munson, W. A. Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(2):82–108, 1933.

Flyvbjerg, B. Five misunderstandings about case-study research. *Qualitative inquiry*, 12(2):219–245, 2006.

Gibson, J. J. *The senses considered as perceptual systems.* Boston: Houghton, Mifflin and Company, 1966.

Gibson, J. J. The theory of affordances. In Shaw, R. and Bransford, J., editors, *Perceiving, acting, and knowing: Toward an ecological psychology*, pp. 67–82: Lawrence Erlbaum, 1977.

Gibson, J. J. *The ecological approach to visual perception.* Boston: Houghton, Mifflin and Company, 1979.

Givan, B. Thelonious Monk's pianism. *Journal of Musicology*, 26(3):404–442, 2009.

Hamilton, A. *Lee Konitz: Conversations on the Improviser's Art*. Ann Arbor: University of Michigan Press, 2007.

Hammersley, M. On the teacher as researcher. *Educational Action Research*, 1(3):425–445, 1993.

Hammersley, M. Is social measurement possible, and is it necessary? In Walford, G., Tucker, E., and Viswanathan, M., editors, *The SAGE Handbook of Measurement*, pp. 409–426: SAGE Publications, 2010.

Haugeland, J. *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press, 1985.

Heider, F. and Simmel, M. An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2):243–259, 1944.

Hendriks-Jansen, H. *Catching ourselves in the act: Situated activity, interactive emergence, evolution, and human thought*. Cambridge: MIT Press, 1996.

Hiller, L. Composing with computers: A progress report. *Computer Music Journal*, 5 (4):7–21, 1981.

Hiller, L. A. and Isaacson, L. M. Musical composition with a high-speed digital computer. *Journal of the Audio Engineering Society*, 6(3):154–160, 1958.

Hofstadter, D. How could a copycat ever be creative? In Dartnall, T., editor, *Artificial Intelligence and Creativity*, pp. 405–424. Dordrecht: Kluwer Academic Publishers, 1994.

Hofstadter, D. R. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York: Basic Books, 1995.

Holliday, A. *Doing and Writing Qualitative Research*. London: SAGE Publications, 2007.

Hsu, W. Strategies for managing timbre and interaction in automatic improvisation systems. *Leonardo Music Journal*, 20:33–39, 2010.

Jackson, M. *Problem frames: analysing and structuring software development problems*. Boston: Addison-Wesley, 2001.

Johnson-Laird, P. N. How jazz musicians improvise. *Music Perception*, 19(3):415–442, 2002.

Kennedy, J. and Eberhart, R. Particle swarm optimization. In *Proceedings of 1995 IEEE International Conference on Neural Networks*, pp. 1942–1948, 1995.

Király, I., Jovanovic, B., Prinz, W., Aschersleben, G., and Gergely, G. The early origins of goal attribution in infancy. *Consciousness and Cognition*, 12(4):752–769, 2003.

Kirsh, D. Today the earwig, tomorrow man? *Artificial intelligence*, 47(1):161–184, 1991.

Koro-Ljungberg, M., Douglas, E., McNeill, N., Therriault, D., and Malcolm, Z. Reconceptualizing and de-centering think-aloud methodology in qualitative research. *Qualitative Research*, in press.

Large, E. W. and Kolen, J. F. Resonance and the perception of musical meter. *Connection science*, 6(2-3):177–208, 1994.

Lehmann, A. C. and Kopiez, R. The difficulty of discerning between composed and improvised music. *Musicae Scientiae*, pp. 113–129, 2010. Special Issue 2010.

Levinson, J. Evaluating musical performance. *Journal of Aesthetic Education*, 21(1): 75–88, 1987.

Levisohn, A. and Pasquier, P. Beatbender: subsumption architecture for autonomous rhythm generation. In *ACM International Conference on Advances in Computer Entertainment Technologies (ACE 2008)*, pp. 51–58, 2008.

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

Lewis, G. Too many notes: Computers, complexity and culture in Voyager. *Leonardo Music Journal*, 10:33–39, 2000.

Lewis, G. Gittin'to know y'all: Improvised music, interculturalism and the racial imagination. *Critical Studies in Improvisation/Études critiques en improvisation*, 1 (1), 2004.

Lewis, G. Mobilitas animi: Improvising technologies, intending chance. *Parallax*, 13(4): 108–122, 2007.

Lewis, G. E. *A power stronger than itself: The AACM and American experimental music*. Chicago: University of Chicago Press, 2008.

Linson, A. Unnecessary constraints: A challenge to some assumptions of digital musical instrument design. In *Proceedings of the International Computer Music Conference (ICMC) 2011, University of Huddersfield, UK*, pp. 421–424, 2011.

Linson, A. The expressive stance: Intentionality, expression, and machine art. *International Journal of Machine Consciousness*, 5(2):195–216, 2013.

Linson, A., Dobbyn, C., and Laney, R. Improvisation without representation: artificial intelligence and music. In *Proceedings of Music, Mind, and Invention: Creativity at the Intersection of Music and Computation (MMI) 2012, The College of New Jersey, USA*, 2012a.

Linson, A., Dobbyn, C., and Laney, R. Critical issues in evaluating freely improvising interactive music systems. In Maher, M., Hammond, K., Pease, A., Pérez y Pérez, R., Ventura, D., and Wiggins, G., editors, *Proceedings of the Third International Conference on Computational Creativity (ICCC)*, pp. 145–149, 2012b.

Linson, A., Dobbyn, C., and Laney, R. Interactive intelligence: Behaviour-based AI, musical HCI and the Turing test. In Müller, V. and Ayesh, A., editors, *Proceedings of the AISB (Society for the Study of Artificial Intelligence and Simulation of Behaviour) and IACAP (International Association for Computing and Philosophy) World Congress, Alan Turing Year 2012, Birmingham, UK*, pp. 16–19, 2012c.

Linson, A., Dobbyn, C., and Laney, R. A parsimonious cognitive architecture for human-computer interactive musical free improvisation. In *Advances in Intelligent Systems and Computing 196: Biologically Inspired Cognitive Architectures 2012*, pp. 219–224. Berlin: Springer, 2013.

Linson, A., Dobbyn, C., Lewis, G., and Laney, R. An artificial agent for collaborative free improvisation. *(submitted)*, in peer review.

Loui, P., Wu, E. H., Wessel, D. L., and Knight, R. T. A generalized mechanism for perception of pitch patterns. *The Journal of Neuroscience*, 29(2):454–459, 2009.

MacDonald, R. and Wilson, G. Constructions of jazz: How jazz musicians present their collaborative musical practice. *Musicae Scientiae*, 10(1):59–83, 2006.

MacDonald, R., Wilson, G., and Miell, D. Improvisation as a creative process within contemporary music. In Hargreaves, D., Miell, D., and Macdonald, R., editors, *Musical Imaginations: Multidisciplinary perspectives on creativity, performance and perception: Multidisciplinary perspectives on creativity, performance and perception*, pp. 242–256. Oxford: Oxford University Press, 2011.

Mataric, M. J. A distributed model for mobile robot environment-learning and navigation. Master's thesis, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1990.

McCorduck, P. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* New York: WH Freeman & Co., 1979.

Menabrea, L. F. and Lovelace, A. K. *Sketch of the Analytical Engine Invented by Charles Babbage, Esq.* London: Taylor and Francis, 1843.

Milne, A., Sethares, W., Laney, R., and Sharp, D. Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, 5(1):1–20, 2011.

Minsky, M. *The society of mind.* New York: Simon and Schuster, 1986.

Mitchell, M. and Hofstadter, D. R. The emergence of understanding in a computer model of concepts and analogy-making. *Physica D: Nonlinear Phenomena*, 42(1): 322–334, 1990.

Monson, I. Doubleness and jazz improvisation: Irony, parody, and ethnomusicology. *Critical Inquiry*, 20(2):283–313, 1994.

Müller, V. Interaction and resistance: The recognition of intentions in new human-computer interaction. In Esposito, A. et al., editors, *Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: COST 2102 International Training School, LNCS*, volume 6456, pp. 1–7. Berlin: Springer, 2011.

Müller, V. Autonomous cognitive systems in real-world environments: Less control, more flexibility and better interaction. *Cognitive Computation*, 4(3):1–4, 2012.

Newell, A. and Simon, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126, 1976.

Newell, A., Shaw, J. C., and Simon, H. A. The processes of creative thinking, 1958.

Nilsson, N. J. Shakey the robot. Technical report, DTIC Document, 1984.

Norman, D. A. Affordance, conventions, and design. *interactions*, 6(3):38–43, 1999.

Pachet, F. The continuator: Musical interaction with style. *Journal of New Music Research*, 32(3):333–341, 2003.

Papadopoulos, G. and Wiggins, G. Ai methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity*, pp. 110–117, 1999.

Pearce, M. and Wiggins, G. Auditory expectation: The information dynamics of music perception and cognition. *Topics in Cognitive Science*, 4(4):625–652, 2012.

Pickering, A. *The cybernetic brain: Sketches of another future.* Chicago: University of Chicago Press, 2010.

Port, R. and Van Gelder, T. *Mind as Motion: Explorations in the Dynamics of Cognition.* Cambridge: MIT Press, 1995.

Port, R. F., Cummins, F., and McAuley, J. D. Naive time, temporal patterns, and human audition. In Port, R. F. and van Gelder, T., editors, *Mind as motion*, pp. 339–371. Cambridge: MIT Press, 1995.

Poulin-Dubois, D. and Shultz, T. R. The development of the understanding of human behavior: From agency to intentionality. In Astington, J. W., Harris, P. L., and Olson, D. R., editors, *Developing Theories of Mind*, pp. 109–125. Cambridge: Cambridge University Press, 1988.

Pressing, J. Improvisation: methods and models. In Sloboda, J., editor, *Generative processes in music*, pp. 129–178. Oxford: Oxford University Press, 1988.

Repp, B. The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14(2):161–184, 1996.

Risset, J.-C. and Van Duyne, S. Real-time performance interaction with a computer-controlled acoustic piano. *Computer music journal*, 20(1):62–75, 1996.

Ritchie, J. and Lewis, J. *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* London: SAGE Publications, 2003.

Rowe, R. Machine listening and composing with Cypher. *Computer Music Journal*, 16 (1):43–63, 1992.

Rowe, R. *Interactive music systems: Machine listening and composition.* Cambridge: MIT Press, 1993.

Sansom, M. J. *Musical meaning: A qualitative investigation of free improvisation.* PhD thesis, University of Sheffield, 1997.

Seddon, F. A. Modes of communication during jazz improvisation. *British Journal of Music Education*, 22(1):47–61, 2005.

Shannon, C. E. Programming a computer for playing chess. *Philosophical magazine*, 41 (314):256–275, 1950.

Simon, H. A. and Newell, A. Computer simulation of human thinking and problem solving. *Monographs of the Society for Research in Child Development*, 27(2):137–150, 1962.

Sklar, E. S. The tribunal of use: Agreement in indefinite constructions. *College Composition and Communication*, 39(4):410–422, 1988.

Sloman, A. Architectural and representational requirements for seeing processes and affordances. In Heinke, D. and Mavritsaki, E., editors, *Computational Modelling in Behavioural Neuroscience: Closing the gap between neurophysiology and behaviour*, pp. 303–331. New York: Psychology Press, 2009.

Smith, H. and Dean, R. *Improvisation, Hypermedia and the Arts since 1945*. Newark: Harwood Academic Publishers, 1997.

Stevens, J. *Search and reflect: A music workshop handbook*. London: Community Music, 1985.

Stock, J. P. Documenting the musical event: observation, participation, representation. In Clarke, E. and Cook, N., editors, *Empirical musicology: Aims, methods, prospects*, pp. 15–34. Oxford: Oxford University Press, 2004.

Suchman, L. *Human–Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press, 2007.

Sudnow, D. *Ways of the Hand: A Rewritten Account*. Cambridge: MIT Press, 2001.

Toop, D. *Haunted Weather: Music, Silence, and Memory*. London: Serpent's Tail, 2004.

Turing, A. M. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

Unemi, T. and Bisig, D. Playing music by conducting Boid agents – a style of interaction in the life with a-life. In *Proceedings of A-Life IX*, pp. 546–550, 2004.

Walter, W. *The living brain*. Oxford: W.W. Norton, 1953.

Walter, W. G. An imitation of life. *Scientific American*, 182(5):42–45, 1950.

Walter, W. G. Frequency analysis. In Brazier, M., editor, *Computer techniques in EEG analysis. Electroencephalography and clinical neurophysiology*, Supplement No. 20, 1960.

Wang, D. and Brown, G. J. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Hoboken: Wiley-IEEE Press, 2006.

Wang, G. *The ChucK Audio Programming Language: A Strongly-timed and On-the-fly Environ/mentality*. PhD thesis, Princeton University, 2008.

Wang, G., Fiebrink, R., and Cook, P. R. Combining analysis and synthesis in the ChucK programming language. In *Proceedings of the International Computer Music Conference*, pp. 35–42, 2007.

Weizenbaum, J. Eliza: a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

Weizenbaum, J. Contextual understanding by computers. *Communications of the ACM*, 10(8):474–480, 1967.

Wengraf, T. *Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods*. London: SAGE Publications, 2001.

Wiener, N. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge: MIT Press, 1948.

Wiggins, G. A. Models of musical similarity. *Musicae Scientiae*, 11(Discussion Forum 4A):315–338, 2007.

Williams, A. and Irani, L. There's methodology in the madness: toward critical HCI ethnography. In *Proceedings of CHI 2010: alt.ernative Methods*, pp. 2725–2734, 2010.

Winograd, T. and Flores, F. *Understanding Computers and Cognition: A New Foundation for Design*. Norwood: Ablex Publishing Corporation, 1986.

Wulfhorst, R. D., Nakayama, L., and Vicari, R. M. A multiagent approach for musical interactive systems. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems*, pp. 584–591. ACM, 2003.

Xenakis, I. *Formalized music: thought and mathematics in composition*. Hillsdale: Pendragon Press, 1992.

Young, M. Identity and intimacy in human–computer improvisation. *Leonardo Music Journal*, 20:97–97, 2010.