



# Open Research Online

---

The Open University's repository of research publications and other research outputs

## A Subsumption Agent for Collaborative Free Improvisation

### Journal Item

How to cite:

Linson, Adam; Dobbyn, Chris; Lewis, George and Laney, Robin (2015). A Subsumption Agent for Collaborative Free Improvisation. *Computer Music Journal*, 39(4) pp. 96–115.

For guidance on citations see [FAQs](#).

© 2015 Massachusetts Institute of Technology

Version: Accepted Manuscript

Link(s) to article on publisher's website:  
[http://dx.doi.org/doi:10.1162/COMJ\\_a00323](http://dx.doi.org/doi:10.1162/COMJ_a00323)

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

« Author preprint of a manuscript accepted for publication in  
Computer Music Journal (MIT Press), to appear 2015 (Vol. 39:4) »

# A Subsumption Agent for Collaborative Free Improvisation

Adam Linson,<sup>\*</sup> Chris Dobbyn,<sup>\*\*</sup> George E. Lewis,<sup>†</sup> and Robin Laney<sup>\*\*</sup>

<sup>\*</sup>Faculty of Music

University of Oxford

Saint Aldate's

Oxford OX1 1DB, UK

adam.linson@music.ox.ac.uk

<sup>\*\*</sup>Faculty of Mathematics, Computing and Technology

Open University

Walton Hall

Milton Keynes MK7 6AA, UK

{chris.dobbyn, robin.laney}@open.ac.uk

<sup>†</sup>Department of Music

Columbia University

621 Dodge Hall, MC1813

2960 Broadway

New York, NY 10027, USA

gl2140@columbia.edu

## Abstract

This paper discusses the design and evaluation of an artificial agent for collaborative musical free improvisation. The agent provides a means to investigate the underpinnings of improvisational interaction. In connection with this general goal, the system is also used here to explore the implementation of a collaborative musical agent using a specific robotics architecture, Subsumption. The architecture of the system is explained, and its evaluation in an empirical study with expert improvisors is discussed. A follow-up study using a second iteration of the system is also presented. The system design and connected studies bring together Subsumption robotics, ecological psychology, and musical improvisation, and contribute to an empirical grounding of an ecological theory of improvisation.

Designers of human-computer interactive systems for musical improvisation have taken diverse approaches to system development. Some of these systems implement an abstract model of the human mind (e.g., Rowe 1992), while others implement a model of the sonic and emotional organizational principles of improvisation, and emergent collective musical behavior (e.g., Lewis 1999). Specific design strategies for producing “weakly to strongly musician-like” interactive performances may cover an extensive range of computational models and techniques including “swarms, recurrent neural networks, and simulations of self-organizing criticality systems such as sand piles and forest fires” (Bown 2011). Significantly, the development of such systems has often been for artistic ends. For those developed primarily to produce output that fulfills aesthetic criteria, a valid design strategy may draw upon whatever technical means are available, rather than be artificially bound to a given model or limited to a particular set of techniques.

The system described below, Adam Linson's *Odessa*, is an artificial agent designed for research into the complex dynamics of the situated psychosocial and embodied cognitive practice of musical improvisation. The design is not intended to faithfully reproduce biological or psychological mechanisms, nor to exhibit humanlike musicality; it does, however, provide a starting point for modeling the sonic-behavioral dynamics of a collaborative musical improviser. Despite its simplified design, the activity of the system is governed by the interactive complexity of real-time, real-world free improvisation. *Odessa* thereby provides a means to investigate the underpinnings of improvisational interaction. In connection with this general goal, the system is also used here to explore the implementation of a collaborative musical agent using a specific robotics architecture (Subsumption).

In the following sections, the musical context relevant to the design of *Odessa* is introduced, a detailed account of the system design is given, and related systems are compared. Then, after a discussion of the initial evaluation, a second design iteration and follow-up study are presented, concluding with a general discussion.

## **Design**

This section presents the design of *Odessa*. It includes some background on the musical context, discusses unstructured input and models, and provides descriptions of the system's streaming output mechanisms and an account of its interactivity. Subsequently, an overview of the system architecture will reveal how its components are integrated. Some examples of module content will also be given. The section concludes with a comparison to related systems.

*Odessa*, implemented in ChuckK (Wang et al. 2007; Wang 2008), functions dynamically as a parsimonious cognitive model from which complex human-interactive musical behavior

emerges. Its computationally lightweight and modular design is a novel application of Rodney Brooks' (1999) subsumption architecture (Subsumption, hereafter), which was originally developed for building mobile robots. In Subsumption terminology, the "competing behaviors" of a system are organized into a network of interactive "layers" (Brooks 1999). Following this idea, and further informed by psychological research, *Odessa's* behavioral layers consist of its basic agency to spontaneously produce output ('Play'), its ability to respond to musical input ('Adapt'), and its ability to resist by disregarding musical input, introducing silence, and initiating endings ('Diverge'). (These layers and related background are discussed in depth further below.)

The design of *Odessa* was inspired by parallels found in two accounts that connect ecological psychology to, respectively, Subsumption robotics (Clark 1997) and musical improvisation (Clarke 2005). By bringing together the latter two, this research in turn contributes to the empirical grounding of an ecological theory of improvisation. Although Brooks has expressed his own narrow view of musical communication, through the suggestion that it is dependent on visual interaction (see Lewis 2007), it is notable that his former student and collaborator Jonathan Connell (whose work led to important revisions in Subsumption) describes their (non-musical) mobile robot as having the ability to "improvise", in this case, describing the navigation of dynamic physical environments such as busy offices (Connell 1989). 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).

## **Musical context: Free improvisation**

Referring to specific historical events that marked the emergence of an international community of free improvisors, Lewis (2004) writes of “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”. This conception underlies the experimental setup of the initial empirical research on *Odessa* (further described in the ‘Methodology’ section). It is assumed that present-day artificial agents do not share the equivalent of human “histories and personalities”. Without consciousness or a capacity for reflection, such agents cannot experience the cultural dimension of improvisation (see Chella and Manzotti 2012). 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; see also Lewis 2000).

Nonetheless, the interactional behavior of improvising can be investigated in terms of the dynamic coupling of human and computer agents. Irrespective of field, any research into human behavior should be expected to account for the rich complexity of an ‘environment’ or ‘situation’, a complexity that is particularly evident in the case of an improvisational musical encounter. It is hoped that *Odessa* can contribute to an understanding of some of the ways in which this complexity arises, and some of the roles it plays in human experience.

## ***Odessa’s* 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. They also have a robust, flexible approach to dealing with unpredictable changes in the environment (Clarke 2005; see also Sudnow 2001).

For Subsumption systems, the ability to “respond quickly to changes in the world” is, in part, achieved by a means of accommodating unstructured input, as opposed to expecting input to conform to an internal model (Brooks 1999, p. 68). Brooks’ (1999) work has shown that his robots function effectively without the use of internal models, that is, without ideal formalizations 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 modeling improvisation. Certain forms of improvisative music have proved amenable to formalized musical description, e.g., using rules for fitting a melody to a chord progression (e.g., Biles 1994). Such musical formalizations, however, are generally regarded as abstractions of an embodied performance tradition that do not necessarily indicate how it is approached by human musicians (see, e.g., Bailey 1980/1993). While there has been some work on formal models of free improvisation, these have typically relied on non-musical formalizations, such as dynamical systems models (e.g., Blackwell and Young 2004; Borgo and Goguen 2005).

Clearly, a non-learning system such as *Odessa* is limited in certain musical respects by its lack of internal models. However, the present research concerns its collaborative role in human–computer interactive free improvisation. It is hypothesized that *Odessa*’s ability to collaborate with experts in this domain is not compromised by these formal musical limitations (see Stevens (1985) for one approach to free improvisation that is neutral with respect to formal musical abilities). Collaboration, in the sense used here, can be

understood as interactive engagement that potentially leads to unanticipated musical outcomes.

## Streaming output

This subsection addresses *Odessa's* note stream formation and decomposition.

### *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. 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 an abstract global timeline. From a traditional perspective, this would seem to present 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: a continuous series of discrete monophonic note streams 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 (synthesized or acoustic).





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

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. In the actual system, such sequences are neither globally quantized nor synchronized to one another in a musical sense (technically, there is hardware-level synchronization due to a shared processor clock, although, in principle, multiple asynchronous processors could be used).

### *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 2004; 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.

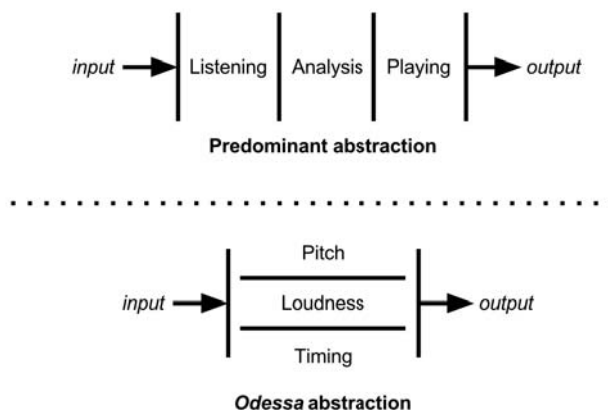


Figure 2. Top: Traditional horizontal decomposition with layers between input and output; Bottom: Subsumption vertical decomposition in which each layer connects input to output (adapted from Brooks (1999), p. 67).

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.<sup>1</sup> In contrast, in other free improvisation systems, one or more of these means are typically used, for example, 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 2004).

With *Odessa*, incoming sound to the soundcard (transduced via microphone or pickup) is

<sup>1</sup>To clarify this point, it is certainly the case that an external observer may interpret the actions of the system in terms of Western musical representations (e.g., semitones, octaves), or in terms of a scientific (mathematical-physical) description (e.g. Hertz, the harmonic series). Indeed, in the ChuckK software used, a number of these abstractions are present for human convenience. However, it is important to stress that at no time are these part of a representational model used *by* the system itself. The system simply extracts the strongest incoming physical vibrations via the transducer and transforms them in isolation of any model (apart from the weak sense of a 'number line' model implicit in incrementing or decrementing values), i.e., it does not make use of any specified relationships between model-internal elements (the usual sense of both a musical and world model).

analyzed in separate, uniparametric dimensions: frequency for pitch approximation, amplitude for loudness approximation, and time between notes. These parameters are concurrently analyzed by dedicated modules (i.e., one for each). Output is formed through an integration of these separate parameter streams (described in the preceding subsection).

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. 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 input analysis. Instead, it uses simple input transformations (described further below).

## **Interactivity**

This subsection gives an account of *Odessa*’s interaction model, its human–computer interactive behavior, and the interaction of its constituent subsystems.

### *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 always be appropriate to think of an independent robot agent as having *users*, it is nonetheless 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 isolated system properties. By design, Subsumption agents “[rely] heavily on the dynamics of their interactions with the world to produce their results” (Brooks 1999, p. 68). As Suchman (2007) points out, 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 (2003), whose *Continuator* is presented as a means to extend an individual’s musical performance capacities. 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 *Voyager*, one of the first computer systems built expressly for this type of music: "there is no built-in hierarchy of human leader / computer follower, no 'veto' buttons, pedals, or cues" (Lewis 1999). This general approach to interaction design is shared by other systems with similar aims, such as those by Blackwell and Young (2004) and Collins (2006), although implementations vary greatly. More generally, this interaction model is opposed to "game-theory models of social interaction that emphasize self-interest", and instead emphasizes coordination, "interdependence", and "mutual control" (Young 2010). The human and computer players function as tightly coupled subsystems, exerting a constant reciprocal influence on one another.

### *Interactive behavior*

Collaborative musical free improvisation is a form of interaction between distinct individuals who collectively negotiate the construction of a musical piece in real time, without anything agreed upon in advance (Bailey 1980/1993, 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, suggests that it understands its actions in relation to its environment, and that it engages in purposive behavior. (Note that attributed or apparent intentionality differs from the philosophical notion of 'intrinsic' intentionality; see Dennett (1987) for a critical discussion.) Research in psychology, discussed below, suggests that a combination of perceptual cues — perceived when observing and interacting with an agent — lead to the attribution of intentionality.

One design goal of *Odessa* 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 behaviors” (Brooks 1999, p. 90), lends itself to the production of such cues when the agent’s behaviors are organized as an interplay of adaptation and resistance, an idea based on insights from psychology (e.g., Poulin-Dubois and Shultz (1988); Csibra (2008); Király et al. (2003); Barrett and Johnson (2003); see also Müller (2011)). Such research 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). Current empirical psychology research on the attribution of intentionality has played an important role in contemporary cognitive modeling (e.g., Baldwin and Baird 2001) and biomedical research (e.g., Castelli et al. 2002).

For the design of *Odessa*, it was hypothesized that the behavioral decomposition into *Play*, *Adapt*, and *Diverge* would serve to produce cues that suggest intentionality. These three levels (or ‘layers’, discussed in more depth below) also reflect the system’s design history, which followed the Subsumption approach of developing and fine-tuning the layers from lowest to highest, with higher levels typically intervening in and modifying the behaviour of lower ones. *Adapt* and *Diverge* form distinct higher-level behaviors of the system, while the basic *Play* mechanism forms the lowest-level behavior. From a design standpoint, adaptation has been interpreted as an adaptation to the musical behavior of the human co-performer, while resistance has been interpreted as producing a divergence from the human behavior, to potentially lead the collaboration in a different musical direction.

### *Layer interaction*

The *Play*, *Adapt*, and *Diverge* behaviors of *Odessa* are separated into Subsumption 'layers' (networks of simple modules), as depicted in Figure 3. Using Brooks' convention, the circles marked 'i' indicate *inhibition* and those marked 's' indicate *suppression* (Figure 3). When data is *inhibited*, the data is blocked from transmission along the line of data flow between modules. When data is *suppressed*, data flowing from one module is replaced by data from a different source module.

In the absence of external (sonic) input from a human co-performer, the *Play* layer generates an independent musical output stream. When external input is detected, the *Adapt* behavior is activated, which results in the output stream adapting to the human co-performer's musical behavior by using pitches, loudness, and timing derived from and closely related to the input source. The design aim here is to give the human performer a sense of *Odessa* cooperating.

However, if this layer remains activated for an extended period, the behavior 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 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 of its internal modules.

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

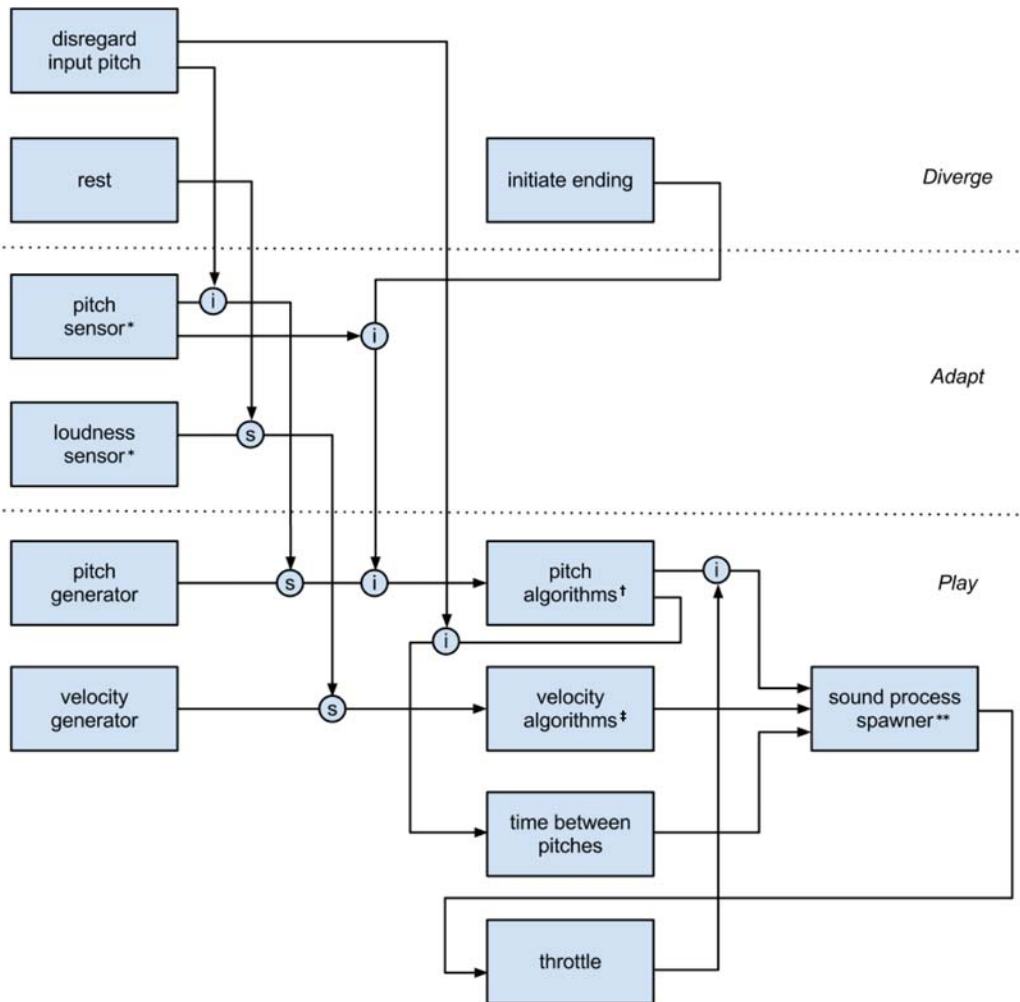


Figure 3. Odessa architecture. Modules are indicated by named boxes. Layers are separated by dotted lines. Solid lines indicate data flow in direction of arrow. Circles marked 'i' indicate inhibition and those marked 's' indicate suppression (see ). \*Receives external audio input. \*\*Transmits external 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.



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.

## **Module simplicity**

The potentially surprising simplicity of *Odessa's* modules, when considered in the context of their practical roles in the system behavior, forms a key strength of this research. Brooks (1999) 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” (p. 3). 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” (p. 3).

One aim of developing an artificial agent for collaborative free improvisation using Subsumption is to demonstrate that complex interactive behavior, subject to evaluation by experts, can emerge from simple interactions between simple modules in a complex environment. Thus, at every instance where a more complex module operation could be substituted, a simple variant has been used instead. The use of simple 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 a complex module were used in place of a simple one, the source of complex interactive musical output could not be exclusively attributed to simple interactions between simple modules.

### *Module examples*

The following are representative examples of module content (see Figure 3), illustrating their simplicity.

**Pitch sensor.** Continuously polls the sonic input signal from human instrumentalist and extracts the strongest frequency values from the spectrum. 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 behavior, 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.

**Pitch algorithms.** Incoming pitch (input from either performer or pitch generator input) is transformed according to an arbitrarily selected operation that 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.

**Time between pitches.** Finds the duration of silence between incoming notes, specifically between a note endpoint and the onset of the next note. This value was empirically determined to be more useful than interonset values or note durations, 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.

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

## **Other Subsumption-related computer music systems**

### *Reactive Accompanist*

Joanna Bryson (1995) was the first to develop a musical agent using Subsumption. Her system, the *Reactive Accompanist* thus relates to *Odessa*, in so far as both are Subsumption agents for music. 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 substantially.

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 modeled 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. These differences are highlighted 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 between the two system designs 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 for a more parsimonious computational implementation, and it does not require any advanced training, which makes it more flexible with respect to performance context (in the sense that, without any training 'abstractions' in the system, there is no potential input that could violate them).

The next point relates to the *Reactive Accompanist's* modeled 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 to pitch class, 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 require strict isochrony. Finally, while the chordal accompaniment of a melody is modeled 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 (this point is elaborated on in the General Discussion).

Apart from these system design differences, for the empirical study of the respective 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). The more elaborate evaluation of *Odessa* is described in the next section.

### *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). And, significantly, with their design, 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.

## **Methodology**

This section describes the empirical evaluation of the system.

### **Experiment context and description**

The empirical evaluation of *Odessa* was designed to maximize 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. However, the experiment was designed to preserve many aspects of a relatively common mode of encounter among the international community of free improvisors: when players who have not performed together, and in some cases not met or heard each other play, engage in real time improvisational musical collaboration.

The participant 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 interviewer (the lead author) 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, the interviewer was able to recognize 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 the role of the interviewer was the sense that, given the (correct) perception that it was the system designer in this role, the question remained as to how critical the participants could be while still feeling tactful and comfortable, in light of the interviewer/designer's potential discomfort during such critique. This raised the issue of the degree to which participants might be holding back more critical responses. Two interrelated strategies (described below) 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).

#### *Format and procedure*

The first strategy was to use unstructured verbal (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 interviewer. 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.

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. 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, 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.

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 gather more specific data (e.g., "You stated that the system responded to your playing 'pretty adequately'. How would you characterize what was *inadequate* about its responses?").

After completing all the individual sessions, the participant data was analyzed for (intra- and intersubject) themes, 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 musical audio recordings of the speaker's improvisations. Additional interrelationships between these complementary data sets were also examined.

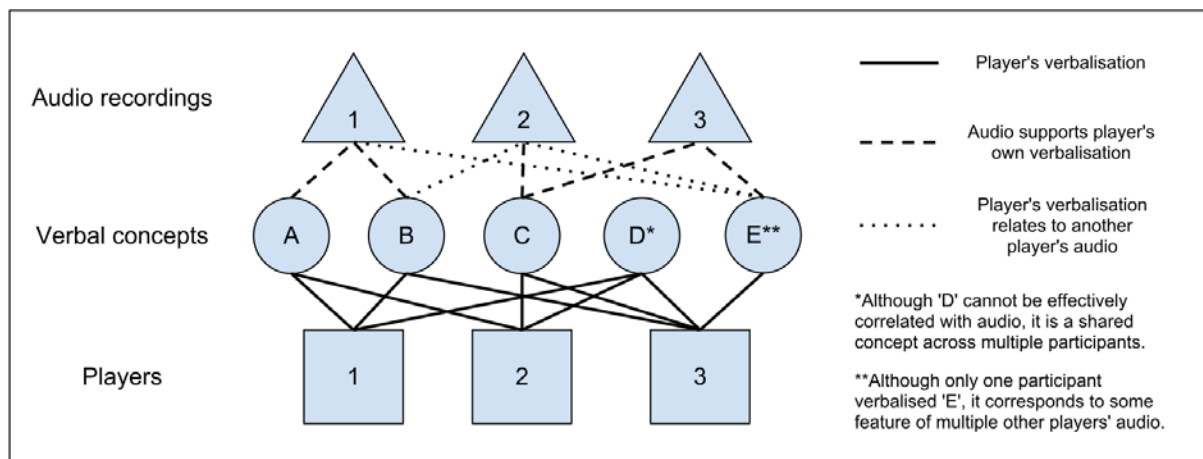


Figure 4. Data relationships.

### Participants

The study consisted of eight experimental 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, vocals. At least four of them 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 by the lead author that later informed one design component of the initial *Odessa* prototype (musical stream

decomposition).

### *Apparatus*

The audio for the cello, guitar, and bassoon was captured using pickups that were 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. Since higher quality directional microphones were not available for the study, having the players use headphones was considered as an alternate 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 decided that the less pristine system response to the player's input, resulting from the occasional intrusions of audio feedback, was preferable to 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 for all studies, although this would have been preferred. For the follow-up study described further below, a Disklavier was used. (Audio of the first iteration of the system using a Disklavier is available here:

<https://soundcloud.com/adamlinson/linson-odessa-iks-goldsmiths/s-ZxqUO>). This performance was not part of the formal study, but was presented by the lead author (on double bass) at the *Interactive Keyboard Symposium*, Goldsmiths, University of London,

2012.)

Discrete pitches and an emulated piano timbre were used in the 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.

## **Results Summary and Discussion**

The case studies suggest that the strategy used to achieve perceived intentionality for collaborative purposes was reasonably effective. To summarize the overall impressions of the studies, six of the eight participants described a process of familiarization and improved collaborative engagement over three duet performances. This is particularly significant given the lack of any machine learning. Two were largely dissatisfied: one participant found no change across performances and another found the standard of performances to have been gradually declining.

A different subset of six players indicated that their take on the machine "anthropomorphized" it, including two who explicitly used that term (or a grammatical variation). Of these six, four struggled to assign the system a gender identity in their

discussion, arising from a preference to refer to the system as “he” or “she” rather than “it”. Notably, the name *Odessa* had not been disclosed to any of the participants. One participant’s tendency to associate the system with humanlike qualities could be discerned through a critical description of the system behaving like “a baby that keeps enjoying its own sound”. In contrast to these perceptions of a potentially lifelike intentional agent, one participant stated “I can’t pretend [I am] playing with another human being”, whereas another simply referred to the system as “the program”. These latter views allay concerns about a potential confirmation bias in the study design, at least to a degree, since the careful use of language while conducting the study resulted in varied participant answers on this point.

Critical feedback from the studies can be categorized under three main headings. The most significant criticisms pertain directly to the architecture. They include problems such as the musical homogeneity of the computer playing, or of the computer output relating either too closely or not closely enough to the human musical input. For instance, one participant used the term “shadowing” to describe its behavior, consistent with a view shared by others that it was at times too closely following what they were doing.

Players also perceived an inability of the system to find (in their words) “common ground” or a “common language”, indicated by statements such as “it doesn’t know how to get into your world” and “I kept looking for something that I could [...] go inside ... and after awhile I [...] stopped looking for this”. These somewhat abstract descriptions were also more directly attributed to the system’s lack of high-level and long-term constructs with comments such as “I felt like it wants to move away from an idea very quickly” and “when I respond to it, it should respond to the fact that I was responding to it”.

A less significant category of critical feedback, though nonetheless relevant to the systems' external technical apparatus, comprised issues that can be clearly linked to extra-architectural issues such as hardware. For example, although there are also psychoacoustic phenomena to consider, a reported dissatisfaction with a perceived lack of changes in loudness (of the system's output) was, at least in part, due to compression in the external amplifier. This was concluded on the basis of audio recordings that bypassed the amplifier, which indicate significant variation in system output amplitude. A perceptibly wide variation in system output loudness was also confirmed when using a Disklavier, which was not part of the initial study. Also, for reasons described above, those using pickups were generally more satisfied with the system's interactions. This suggests the need for a stricter approach to feedback prevention in the apparatus setup for future studies.

The third category of criticism pertains to deliberately imposed experimental constraints. These are also of interest, as they give some indication of the performers' general inclinations. For example, while the emphasis on the performative over the sonic dimension in the experimental design appeared to be generally successful, as indicated by comments such as "it felt quite organic to me to improvise with a piano", some participants found the amplified software synthesis to be problematic for effective acoustic musical interaction. This suggests a potential refinement by using an electromechanically controlled acoustic piano (e.g., a Disklavier), which was part of the follow-up study described below.

On the other hand, a problem that arose directly from the use of piano sounds was that, in some cases, it led to an expectation of humanlike piano competency. This was problematic in the sense that the system did not take pianistic skill into account; for the purposes of the experiment, there was a mere addition of a hard-coded upper and lower bound to constrain output within the piano pitch range. The system was thus unable to

actively form chords conforming to socialized expectations and unable to match the usual traversal patterns of human hands, although idiosyncratic vertical and polyphonic structures emerged from its output. These limitations were a deliberate design choice based on the aims of the study; from a technical standpoint, if the aim were to emulate human piano competence, it would be straightforward to use known probabilities for note transitions and concurrence.

Several participants were also disappointed by the absence of timbres such as those made by directly manipulating the inside of an acoustic piano, and some expressed a desire to interact with more radical electronic timbres. These considerations underscore the significance of topics such as embodiment and culturally situated aesthetic sensibilities to music cognition research.

## **Follow-up Study**

Based on participant feedback from the first study and additional theoretical considerations, a second iteration of *Odessa* was developed and tested.

### **Second iteration design**

To improve *Odessa's* interaction ability, further emphasis was given to its ability to adapt to changes in musical context. Its lack of contextually significant adaptation was identified as a shortcoming in the previous round of participant feedback. Port, Cummins, and McAuley (1995) discuss entrainment as a general basis of adaptation and pattern recognition in an ecological context. 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, 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, for the second iteration of *Odessa*, a module was added with a virtual oscillator for each discrete pitch, which would get "excited" by incoming pitches (i.e., would entrain to them), and gradually decay. With this, 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. 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. Significantly, this iteration still avoids a 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 (based in part on Clarke (2005) and Clark (1997)). On the other hand, the lack of pattern recognition is a significant limitation for which appropriate solutions will be considered in future research.



## 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 synthesized 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 synthesized to acoustic piano.

The follow-up study used two participants from the first study, renowned improvisors Evan Parker (soprano saxophone) and John Russell (guitar), to facilitate a comparison across studies. The participants were, at this point, aware of the system design and the general experimental approach and initial results, as they had been sent documentation of the system design and first study. The design extension, however, was not disclosed to the participants prior to the study. Overall, the follow-up study focused on different issues. It 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 follow-up study participants (Parker and Russell) have also kindly granted their permission to use the musical audio from the follow-up study as supplementary material:

<<https://soundcloud.com/adamlinson/odessa-2nd-study-parker-1/s-4imb6>>.

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, they took part in a group interview that allowed some shared

themes to be explored.

## **Results summary and discussion**

For the second iteration of the system, the participant feedback suggests that, positively, the system exhibited a cohesive, unified identity, and that it was typically responsive, playful and more capable of context sensitivity. However, it still lacked the ability to recognize human responses as such, which meant that it could not engage in a dramaturgical escalation of musical ideas. Its lack of emotion was also found to be significant. It is interesting to note that the lack of emotion was not identified prior to the second iteration, perhaps because the system's improved adaptive capability increased expectations of its behavior.

Concerning *Odessa's* cohesive, unified identity in the second iteration, for instance, the interviewer (the lead author) asked if it seemed that there was one person sitting at the piano the whole time, or if it might have seemed that someone left and then someone else came along. Parker (EP) responded, "no, it didn't feel like that to me, it felt like one person," and Russell (JR) gave a similar response: "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, still resulted in a coherent identity for the system, and was perhaps enhanced by it. In related comments, JR stated that "it felt like there was a kind of broader identity than just a sort of immediate kind of stimulus-response thing in the present"; "it didn't feel like it was a kind of random response. You know, there was something under the hood that was- [...] you know, like personality or something to it".

At one point, JR mentioned that the system had an "incessant" quality, and in a follow-up question, he was asked if this meant it was more like playing along with a recording. He responded that this was not at all the case: "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. [...] I mean trying to improvise to pre-recorded stuff is awful". He concluded that "the program is very playful, that's the point I'm trying to make. It's very playful, and that's fun ... the program responds enough to be playful" (JR).

In the trio context, EP found that the experience of performing with the system was helped by the interplay with another human improviser (JR). As EP stated, "I quite enjoyed the piano playing in that [trio] context. And, as I say, I found it easier to deal with because I knew [JR] 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 [JR] and see what the piano would do." 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 JR, and that this was able to facilitate an experience more similar to experiences with human players.

Perhaps due to the improvements in the second iteration of the system, certain inherent limitations were brought into sharper relief: "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" (EP). It was also noted that "it doesn't have much sense of dramaturgy", "total form" or "emotional significance" (EP), and "it's not got a sense of irony that a human would have" (JR).

## General Discussion

It seems reasonable to conclude that at least some of *Odessa's* limitations can be viewed as a result of the design premises, setting aside the specific implementation. In particular, a substantial addition to the design would be required to overcome its presently identified limitations, possibly compromising the fundamentals of a Subsumption design. Ultimately, the tendency of the system to return to a narrow range of behavior, which can be conceptualized as its tendency to homeostasis (Ashby 1952/1960), has been experienced as a drawback in the aesthetic-social realm of free improvisation.

However, it is possible to consider some interesting relationships between *following*, pattern recognition, entrainment, and ecological psychology (in particular, the theory of affordances), with respect to the present study. 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 synchronizes, or entrains, a listener's internal processing mechanisms". Their solution for modeling this phenomenon is to use a network of dynamical systems that can "self-organize 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).

In designing *Odessa*, it was assumed that a performer playing along with static (e.g. pre-recorded) source material would not have the experience of a collaborative performance. At the other end of the spectrum, it was known that computationally intricate systems could produce sophisticated humanlike improvisational behavior (cf. Braasch et al. 2012, Van Nort et al. 2013). The research presented above sought to investigate if a fundamentally simple system could be dynamic enough to engage an expert human improviser, or if it would remain closer to the experience of interacting with a static source.

The hypothesized viability of the 'simple system' approach rested on the premise of using basic cues to induce the performer's attribution of intentionality to the agent, implemented using Subsumption. For *Odessa*, the mechanisms producing psychological cues for the perception of intentional agency seem to be effective in establishing it as a legitimately collaborative partner. Even with its limited musicality, it also seems to produce the right cues for facilitating the general behavior of collaborative free improvisation. However, it is clear that its lack of understanding of musical significance, of long-term musical structure, of emotion and dramaturgy, are critical limitations.

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 improviser, 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. It is improbable that an artificial agent could exhibit this level of understanding without a vastly more complex apparatus, such as one with deep similarities to human biology.

## Conclusions and Future Work

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 behavior of a collaborative improviser. This 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. It has been shown that a parsimonious Subsumption system can achieve a complex and robust musical interaction that goes a remarkable distance towards 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 behavioral cues perceived in real time, can lead to the attribution of intentional agency. Furthermore, the fact that the musical behavior 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.

Continued research on *Odessa* will focus on two main areas and their interrelationships, namely, both individual and social aspects of improvised music-making, especially from an ecological perspective. At the individual level, neural models of memory, attention and inference in musical improvisation will be investigated. At the social level, questions of co-creativity, distributed cognition, and related topics will be explored.

In the diverse field of interactive improvisational music systems, the present research is both a continuation of earlier work by others and a starting point for a new direction. Current insights from the research on *Odessa* can be applied to other interactive improvisation systems, for a variety of both aesthetic and scientific research purposes. It is hoped that future work with the system can also contribute to other areas of scientific inquiry and unique cultural pursuits.

## **Acknowledgments**

Many thanks to the anonymous reviewers and the editors, Doug Keislar and Doug Van Nort, for their valuable feedback on the manuscript, and to the participants for their involvement in the studies. Adam Linson also thanks Antonio Chella, Rebecca Fiebrink,

Allan Jones, and Alistair Zaldua for opportunities to present some of this work, and Martyn Hammersley, Michael A. Jackson, Geraint Wiggins, Roger Dean, Andrew Brown, Neal Farwell, and Michal Mukawa for conversations that led to improvements in this paper.

## References

- Angelis, V., S. Holland, P. J. Upton, and M. Clayton. 2013. "Testing a computational model of rhythm perception using polyrhythmic stimuli." *Journal of New Music Research* 42(1):47–60.
- Ashby, W. R. 1952/1960. *Design for a brain: The origin of adaptive behaviour*. Chapman & Hall.
- Assayag, G., G. Bloch, M. Chemillier, A. Cont, and S. Dubnov. 2006. "Omax brothers: A dynamic topology of agents for improvisation learning." In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*. pp. 125–132.
- Bailey, D. 1980/1993. *Improvisation: its nature and practice in music*. Cambridge: Da Capo Press.
- Baldwin, D., and J. Baird. 2001. "Discerning intentions in dynamic human action." *Trends in cognitive sciences* 5(4):171–178.
- Barrett, J. L., and A. H. Johnson. 2003. "The role of control in attributing intentional agency to inanimate objects." *Journal of Cognition and Culture* 3(3):208–217.
- Biles, J. 1994. "GenJam: A genetic algorithm for generating jazz solos." In *Proceedings of the International Computer Music Conference*. pp. 131–137.



- Blackwell, T., and M. Young. 2004. "Swarm Granulator." In G. Raidl, et al., (editors) *Applications of Evolutionary Computing: EvoWorkshops 2004, LNCS*, volume 3005. Springer, pp. 399–408.
- Borgo, D., and J. Goguen. 2005. "Rivers of consciousness: The nonlinear dynamics of free jazz." In L. Fisher, (editor) *Jazz Research Proceedings Yearbook*. pp. 46–58.
- Bown, O. 2011. "Experiments in modular design for the creative composition of live algorithms." *Computer Music Journal* 35(3):73–85.
- Braasch, J., D. Van Nort, P. Oliveros, S. Bringsjord, N. S. Govindarajulu, C. Kuebler, and A. Parks. 2012. "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*.
- Brooks, R. A. 1999. *Cambrian intelligence: The early history of the new AI*. Cambridge: MIT Press.
- Bryson, J. 1995. "The Reactive Accompanist." Master's thesis, Department of Artificial Intelligence, University of Edinburgh.
- Castelli, F., C. Frith, F. Happé, and U. Frith. 2002. "Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes." *Brain* 125(8):1839–1849.
- Chella, A., and R. Manzotti. 2012. "Jazz and Machine Consciousness: Towards a New Turing Test." In V. Müller, and A. Ayesh, (editors) *Revisiting Turing and his test: Comprehensiveness, qualia, and the real world (AISB/IACAP symposium, Alan Turing Year 2012)*. pp. 49–53.
- Clark, A. 1997. *Being There: Putting Brain, Body, and World Together Again*. Cambridge: MIT Press.

- Clarke, E. 2005. *Ways Of listening: An ecological approach to the perception of musical meaning*. Oxford: Oxford University Press.
- Clayton, M., R. Sager, and U. Will. 2005. "In time with the music: The concept of entrainment and its significance for ethnomusicology." In *European meetings in ethnomusicology*, volume 11. pp. 3–142.
- Collins, N. 2006. "Towards Autonomous Agents for Live Computer Music: Real-time Machine Listening and Interactive Music Systems." Ph.D. thesis, University of Cambridge.
- Conklin, D., and I. H. Witten. 1995. "Multiple viewpoint systems for music prediction." *Journal of New Music Research* 24(1):51–73.
- Connell, J. 1989. "A colony architecture for an artificial creature." Technical report, DTIC Document.
- Cope, D. 2005. *Computer Models of Musical Creativity*. Cambridge: MIT Press.
- Corkill, D. D. 1991. "Blackboard systems." *AI expert* 6(9):40–47.
- Csibra, G. 2008. "Goal attribution to inanimate agents by 6.5-month-old infants." *Cognition* 107(2):705–717.
- Dannenberg, R. B. 1985. "An On-Line Algorithm for Real-Time Accompaniment." In *Proceedings of the 1984 International Computer Music Conference, Paris*. pp. 193–198.
- Dannenberg, R. B., Z. Jin, N. E. Gold, O.-E. Sandu, P. N. Palliyaguru, A. Robertson, and A. Stark. 2013. "Human-Computer Music Performance: From Synchronized Accompaniment to Musical Partner." In *Proceedings of Sound and Music Computing Conference 2013, Stockholm, Sweden*. pp. 136–141.

- Davidson, J. W. 2004. "Music as social behavior." In E. Clarke, and N. Cook, (editors) *Empirical musicology: Aims, methods, prospects*. Oxford: Oxford University Press, pp. 57–75.
- Dennett, D. 1987. *The Intentional Stance*. Cambridge: MIT Press.
- Doffman, M. 2009. "Making It Groove! Entrainment, Participation and Discrepancy in the 'Conversation' of a Jazz Trio." *Language & History* 52(1):130–147.
- Ensmenger, N. 2011. "Is chess the drosophila of artificial intelligence? A social history of an algorithm." *Social Studies of Science* 42(1):5–30.
- Heider, F., and M. Simmel. 1944. "An experimental study of apparent behavior." *The American Journal of Psychology* 57(2):243–259.
- Hsu, W. 2010. "Strategies for managing timbre and interaction in automatic improvisation systems." *Leonardo Music Journal* 20:33–39.
- Jackson, M. 2001. *Problem frames: analysing and structuring software development problems*. Boston: Addison-Wesley.
- Király, I., B. Jovanovic, W. Prinz, G. Aschersleben, and G. Gergely. 2003. "The early origins of goal attribution in infancy." *Consciousness and Cognition* 12(4):752–769.
- Koro-Ljungberg, M., E. Douglas, N. McNeill, D. Therriault, and Z. Malcolm. in press. "Re-conceptualizing and de-centering think-aloud methodology in qualitative research." *Qualitative Research* .
- Large, E. W., and J. F. Kolen. 1994. "Resonance and the perception of musical meter." *Connection science* 6(2-3):177–208.
- Levisohn, A., and P. Pasquier. 2008. "BeatBender: subsumption architecture for autonomous rhythm generation." In *ACM International Conference on Advances in Computer Entertainment Technologies (ACE 2008)*. pp. 51–58.

- Lewis, G. 1999. "Interacting with latter-day musical automata." *Contemporary Music Review* 18(3):99–112.
- Lewis, G. 2000. "Too many notes: Computers, complexity and culture in voyager." *Leonardo Music Journal* 10:33–39.
- Lewis, G. 2004. "Gittin'to Know Y'all: Improvised Music, Interculturalism and the Racial Imagination." *Critical Studies in Improvisation/Études critiques en improvisation* 1(1).
- Lewis, G. 2007. "Mobilitas animi: Improvising technologies, intending chance." *Parallax* 13(4):108–122.
- MacDonald, R., and G. Wilson. 2006. "Constructions of jazz: How jazz musicians present their collaborative musical practice." *Musicae Scientiae* 10(1):59–83.
- Müller, V. 2011. "Interaction and resistance: The recognition of intentions in new human-computer interaction." In A. Esposito, et al., (editors) *Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: COST 2102 International Training School, LNCS*, volume 6456. Springer, pp. 1–7.
- Müller, V. 2012. "Autonomous Cognitive Systems in Real-World Environments: Less Control, More Flexibility and Better Interaction." *Cognitive Computation* 4(3):1–4.
- Pachet, F. 2003. "The continuator: Musical interaction with style." *Journal of New Music Research* 32(3):333–341.
- Port, R. F., F. Cummins, and J. D. McAuley. 1995. "Naive time, temporal patterns, and human audition." In R. F. Port, and T. van Gelder, (editors) *Mind as motion*. Cambridge: MIT Press, pp. 339–371.
- Poulin-Dubois, D., and T. R. Shultz. 1988. "The development of the understanding of human behavior: From agency to intentionality." In J. W. Astington, P. L. Harris, and

- D. R. Olson, (editors) *Developing Theories of Mind*. Cambridge: Cambridge University Press, pp. 109–125.
- Rowe, R. 1992. “Machine listening and composing with Cypher.” *Computer Music Journal* 16(1):43–63.
- Sansom, M. J. 1997. “Musical meaning: A qualitative investigation of free improvisation.” Ph.D. thesis, University of Sheffield.
- Stevens, J. 1985. *Search and reflect: A music workshop handbook*. Community Music.
- Stock, J. P. 2004. “Documenting the musical event: observation, participation, representation.” In E. Clarke, and N. Cook, (editors) *Empirical musicology: Aims, methods, prospects*. Oxford: Oxford University Press, pp. 15–34.
- Suchman, L. 2007. *Human–Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.
- Sudnow, D. 2001. *Ways of the Hand: A Rewritten Account*. Cambridge: MIT Press.
- Van Nort, D., P. Oliveros, and J. Braasch. 2013. “Electro/Acoustic Improvisation and Deeply Listening Machines.” *Journal of New Music Research* 42(4):303–324.
- Wang, G. 2008. “The Chuck Audio Programming Language: A Strongly-timed and On-the-fly Environ/mentality.” Ph.D. thesis, Princeton University.
- Wang, G., R. Fiebrink, and P. R. Cook. 2007. “Combining analysis and synthesis in the Chuck programming language.” In *Proceedings of the International Computer Music Conference*. pp. 35–42.
- Wulfhorst, R. D., L. Nakayama, and R. M. Vicari. 2003. “A multiagent approach for musical interactive systems.” In *Proceedings of the second international joint conference on autonomous agents and multiagent systems*. ACM, pp. 584–591.

Xenakis, I. 1992. *Formalized music: thought and mathematics in composition*. Hillsdale:  
Pendragon Press.

Young, M. 2010. "Identity and Intimacy in Human-Computer Improvisation." *Leonardo  
Music Journal* 20:97-97.