Improvisation without Representation

Artificial Intelligence and Music

Adam Linson, Chris Dobbyn and Robin Laney
Faculty of Mathematics, Computing and Technology
Department of Computing
The Open University
Milton Keynes, UK
{a.linson, c.h.dobbyn, r.c.laney}@open.ac.uk

Abstract—**The title is meant in two senses: both as a reference to the fact that contemporary freely improvised music is under-represented in the study of AI and music; and, as a reference to the behavioral robotics research of Rodney Brooks, who famously stated that in place of computational representation, "the world is its own best model". Brooks cites the theoretical work of Marvin Minsky as a source for his approach to designing intelligent robots. Both Minsky and Brooks describe decentralized, agent-based systems that could arguably be regarded as improvisational, in that they are designed to cope with a dynamic external world. Though neither author explicitly identifies improvisation as a key aim of their approach, some AI systems are, in fact, explicitly designed to engage in improvisation. One example is George Lewis' *Voyager*, a human-computer interactive musical improvisation system that negotiates and contributes to complex sonic environments in real time. This paper describes a new system, Adam Linson's *Odessa*, that interacts with human musicians to perform freely improvised music, following Lewis' interaction model. *Odessa* is designed using Brooks' subsumption architecture, thus bypassing central representation and control in favor of a decentralized and environmentally situated approach.**

Artificial intelligence (AI); music; improvisation; human-computer interaction (HCI); interactive systems; robotics; subsumption architecture

I. Introduction

The title is meant in two senses: both as a reference to the fact that contemporary freely improvised music is under-represented in the study of AI and music; and, as a reference to the behavioral robotics research of Rodney Brooks, who famously stated that in place of computational representation, "the world is its own best model". Brooks cites the theoretical work of Marvin Minsky as a source for his approach to designing intelligent robots [1] (see also [2]). Both Minsky and Brooks describe decentralized, agent-based systems that could arguably be regarded as improvisational, in that they are designed to cope with a dynamic external world (see, for example, [3], [1]).

Though neither author explicitly identifies improvisation as a key aim of their approach, some AI systems are, in fact, explicitly designed to engage in improvisation. One example is George Lewis' *Voyager*, a human-computer interactive musical improvisation system[1] that negotiates and contributes to complex sonic environments in real time [5]. Lewis' interaction design is rooted in the musical practice of free improvisation, and his system successfully interacts with humans in a variety of real-world performance settings; it is also widely referred to in the academic literature on interactive music systems (see, for example, [6], [7]).

Overall, however, very few interactive music systems have been designed for free improvisation, despite the fact that it is a contemporary musical practice, widely documented in audio recordings and academic research (see, for example, [8], [9]). From a system design perspective, the fact that freely improvised music is performed without notation or pre-arranged organization poses a significant difficulty: how to most effectively represent its musical content? Systems for generating style-based improvisational music (e.g., [10], [11]) rely on a particular set of salient aspects of the music, similar to those derived from composed music (such as melodic and harmonic sequences). However, when designing systems for music without an explicit rule-based framework—music such as free improvisation—a different set of salient aspects of the music are relevant to the system behavior, both for input analysis and output generation. This paper describes a new system, Adam Linson's *Odessa*, that interacts with human musicians to perform freely improvised music, following Lewis' interaction model. *Odessa* is designed using the subsumption techniques pioneered by Brooks, thus bypassing central representation and control in favor of a decentralized and environmentally situated approach.

II. Beyond Musical Notation

The accounts of composed music given by Deliege, et al. [12] and Minsky [13] recognize a set of salient aspects in common between composed and freely improvised music. As Smith and Dean state, "there is no absolute opposition between improvisation and composition, only a gradient of creative endeavor from pure improvisation to complete composition"

1 This excerpt of a 1985 interview with George Lewis may be of interest in the present context: "I remember talking to Marvin Minsky and Maryanne Amacher once in SoHo [...] and I said I was interested in buying a computer and building an interactive improvisation system with it. [...] Everyone at the table seemed to think that it was a good idea that should be tried" [4].

[14]. Deliege, et al., in their studies on listener perception of composed (notated) music, 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" [12]. Their account of the salience of these "subjective" cues resonates with an account of musical composition given by Minsky:

> Music need not, of course, confirm each listener's every expectation; each plot demands some novelty. Whatever the intent, control is required or novelty will turn to nonsense. [...] Composers can have different goals: to calm and soothe, surprise and shock, tell tales, stage scenes, teach new things, or tear down prior arts. [...] When expectations are confirmed too often, the style may seem dull. [...] Each musical artist must forecast and predirect the listener's fixations to draw attention *here* and distract it from *there* [13] (emphasis in original).

Following the terminology of Deliege, et al., a performer engaged in freely improvised music may well be interacting with what might constitute objective cues, but arguably, the more specialized skill at work is the artful management of the "psycho-dynamic functions" of subjective cues; this entails, at least in part, the active refocusing of the listener's attention through continuous adjustments in musical output (see [4]). In improvisation, "meaning is created in [real-time] performance as the collision or negotiation of different sets of meaning: [...] that which individual performers perceive and/or mediate; that which the audience expects and that which they receive; and so on" [15] (cited in [14]). These "different sets of meaning", or "frames", will be revisited below.

III. Embodiment, Robotics And Improvisation

In contrast to the mastery of a formalized set of rules, performer skill in freely improvised music can be described in terms of embodied expertise, similar to driving a car: "the expert driver feels when slowing down is required and 'knows' how to perform the action without calculating and comparing alternatives". More generally, "when one achieves [embodied] expertise one can *immediately* both perceive the nature of the situation and 'know' what must be done" [16] (emphasis added; see also [17]). For Brooks, embodiment is a distinctive feature of his style of robotics. His robots' "actions are part of a dynamic with the world and have *immediate* feedback on their own sensations" [1] (emphasis added). Regarding immediacy, consider Minsky's account of perception and how one deals with surprises. He describes taking in a visual scene in a single glance:

> seeing is really an extended process. It takes time to fill in details, collect evidence, make conjectures, test, deduce, and interpret in ways that depend on our knowledge, expectations and goals. Wrong first impressions have to be revised. Nevertheless, all this proceeds so quickly and smoothly that it seems to demand a special explanation [3].

His description of an extended process that takes place in an instant is highly suggestive of the complex processes at work in expert improvisation, both musical and otherwise. Though explicit references to improvisation are scarce in the literature on robotics, the term 'improvise' does appear in a report on mobile robotics authored by Jonathan Connell (a student of Brooks and Minsky), in precisely the sense of rapidly negotiating a complex and highly dynamic environment [18] (curiously, two of the three occurrences of the term are in quotation marks, though it is apparently used consistently). Connell identifies the "high degree of similarity" to Brooks' approach, but favors some ideas more closely aligned, "at least in spirit", with Minsky. Connell points out that his approach is in essence a "refinement" of Brooks' architecture (described below), and that Brooks ultimately adopted the ideas developed in his report.

IV. Application to Computer Music Systems

Music-notation-based representations are typical in music software, but they carry the risk of over-simplifying the musical input and output. Interactive music systems with over-simplified representations, once deployed in real musical environments, may fail to cope with the encountered level of complexity (see, for example, [19]). Brooks offers an alternate strategy, in that his robots are not programmed to computationally reconstruct a complete representation of the outside world. His "subsumption architecture" uses simple agents, with no rich representations of the external environment. There is no central model of the world, and no central locus of control. Instead, sensors provide data to modules designed for immediate input-output reactions. The modules are then integrated into layers of an interactive network from which complex behavior can emerge [1].

Brooks' approach can be extended to the design of freely improvising music systems, to avoid the difficulty of symbolically representing, for example, polyphonically overlapping phrase structures, which pose significant problems for computational analysis [20]. And though more powerful tools may render polyphony tractable, a different type of problem for computational analysis lies in the fact that in a given free improvisation, "multiple referent frames can occur" among participants [14]—or, in other words, multiple simultaneous subjective cues [12]. "Thus," Smith and Dean conclude, "a multiplicity of semiotic frames can be continually merging and disrupting during a 'free' [...] improvisation" [14]. Although Smith and Dean introduce the concept of semiotics, their notion of frames is reminiscent of the description given in Minsky:

> for non-visual kinds of frames, the differences between the frames of a system can represent [...] changes in conceptual viewpoint. Different frames of a system share the same terminals; this [...] makes it possible to coordinate information gathered from different viewpoints [3].

Minsky's frame approach is taken up more concretely in terms of computational musical analysis and generation by Conklin and Witten with their notion of viewpoint decomposition [21] (see also [22]). Viewpoints are independent abstractions for "expressing events in a sequence" in terms of a single parameter of an event's "internal structure" (e.g., pitches, intervals, durations). To form complete musical sequences, a set of individual abstractions are recombined into linked viewpoints, analogous to Minsky's notion of frame-systems [3], [13]. It is worth noting that this approach is not without precedent in the arts. In a 1962 paper, composer Milton Babbitt 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 [23]. And as early as 1939, filmmaker Sergei Eisenstein describes an approach 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" [24] (emphasis in original).

V. Current Research: Adam Linson's *Odessa*

Odessa, a new system designed by Adam Linson, draws on concepts from both the subsumption architecture and musical viewpoint decomposition. It is currently under active development and has already shown promise in preliminary studies. The system is designed for freely improvised music with a human duet partner, conceived of as an "an open-ended and performative interplay between [human and computer] agents that are not capable of dominating each other" [25]. This follows the interaction model of *Voyager*, which involves, as Lewis describes, "two parallel streams of music generation, that of the computer and that of the human, each informed by the other's music—an improvisational, nonhierarchical, subject-subject model of discourse" [5]. One design goal of *Odessa*, already achieved in the prototype, is to give the human performer the sense that the system is collaboratively co-creating the total musical output, without limiting the complexity of the human performance. More extensive studies are planned with a future iteration.

The system is constructed using simple modules that are combined into three layers. The lowest-level layer is responsible for playing music. It consists of two generators, one for pitch and one for velocity, two modules for algorithmic operations on incoming pitch and velocity values, and another for measuring duration between pitch signals. These modules act in terms of three viewpoints, though they do not use prediction or learning, nor do they rely on state. Another module links the viewpoints, thus merging pitch, velocity and duration data in a spawned process that drives sound production (as a virtual piano via speaker output, or, as an electromechanically-controlled acoustic piano). A final module for this layer acts as a throttle to regulate the number of parallel sound-producing processes. Algorithmic operations on incoming values include generating a set of neighboring values and filtering elements from repetitive sequences. The middle layer is responsible for (musically) adapting to input, where a standard audio analog-to-digital converter is used as a sensor. In subsumption terms, the input data "suppresses" the data from the independent generators in the lowest layer. The third layer adds behaviors that diverge from the system input. This layer could be thought of as adding "lifelike" characteristics—in Grey Walter's terms, "to maintain a lifelike attitude it must be adventurous, but not reckless" [26]—although the system does not aim to imitate life, nor is it designed to be mistaken for a human performer (in this respect, the speed and density of the system output may exceed the physiological limits of human acoustic performance). Here, the third-level behaviors serve both to engage and to stimulate or provoke the human interlocutor. This layer includes separate modules for occasionally disregarding input and introducing silence. It also includes a module to initiate ending the performance, which itself can be overridden ("inhibited") by further input.

VI. Conclusion

Among other benefits, the approach described above provides a means of achieving an important balance: it prevents system output that is apparently unrelated to the human partner's contribution, but it also avoids a complete mirroring of the human performance. With interactive systems for freely improvised music, as with mobile robots, a full account of expected input is impossible to predict at design time. Any preformed ideas that an improvisor might have going into a freely improvised performance exist within a sociocultural framework (see [5], [14]), but are analogous to the rules of physical-spatial interaction that are carefully engineered into robots built using the subsumption architecture. Once the fixed system-internal relations are settled upon through empirical testing, both musical improvisation and mobile robotic systems can be deployed in their respective dynamic environments [1] (see also [4]). Ultimately, foregoing central symbolic representation in the design of interactive freely improvising music systems can aid in the achievement of rapid adaptation to ever-changing complex musical environments. This approach can contribute to the development of a wide variety of systems for nonhierarchical human-computer collaboration, in aesthetic realms as well as pragmatic ones.

Acknowledgment

Adam Linson thanks Michael A. Jackson, George Lewis and Raymond MacDonald for sharing their expertise during conversations about topics covered in this paper.

References

[1] R. Brooks, *Cambrian Intelligence: The early history of the new AI*. MIT Press, 1999.

[2] M. Minsky, *The Society of Mind*. Simon & Schuster, 1986.

[3] M. Minsky, "A framework for representing knowledge," in *The Psychology of Computer Vision*, P. Winston, Ed. McGraw-Hill, 1975.

[4] C. Roads, "Improvisation with George Lewis," in *Composers and the Computer*, C. Roads, Ed. William Kaufman, Inc. 1985.

[5] G. Lewis, "Interacting with latter-day musical automata," *Contemporary Music Review*, 18:3, 99-112, 1999.

[6] R. Rowe, *Machine Musicianship*. MIT Press, 2001.

[7] R. T. Dean, *Hyperimprovisation: computer-interactive sound improvisation*. A-R Editions, 2003.

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

[9] D. Borgo, *Sync or swarm: improvising music in a complex age*. Continuum International Publishing Group, 2005.

[10] F. Pachet, "Interacting with a Musical Learning System: The Continuator," in *Proc. of music and artificial intelligence: Second International Conference* (ICMAI 2002), Edinburgh, Scotland, UK, September 12-14, 2002.

[11] J. Biles, "GenJam in perspective: A tentative taxonomy for GA music and art systems," *Leonardo* 36, 1: 43:45, 2003.

[12] I. Deliège, M. Mélen, D. Stammers, and I. Cross, "Musical Schemata in Real-Time Listening to a Piece of Music," *Music Perception* 14(2): 117–60, 1997.

[13] M. Minsky, "Music, Mind, and Meaning," *Computer Music Journal*, vol. 5, no. 3, pp. 28-44, Autumn 1981.

[14] H. Smith and R. T. Dean, *Improvisation, Hypermedia and the Arts since 1945*. Routledge, 1997.

[15] A. Frost and R. Yarrow, *Improvisation in drama*. Macmillan, 1990.

[16] B. Flyvbjerg, "Sustaining non-rationalized Practices: Body-Mind, Power and Situational Ethics," *PRAXIS International*, Issue 1, pp. 92-113, 1991.

[17] H. L. Dreyfus and S. E. Dreyfus, *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Blackwell, 1986.

[18] J. H. Connell, "A colony architecture for an artificial creature," DTIC Document, 1989.

[19] R. D. Wulfhorst, L. Nakayama, and R. M. Vicari, "A multiagent approach for musical interactive systems," in *Proceedings of the second international joint conference on autonomous agents and multiagent systems*: 584-591, 2003.

[20] J. Pressing, "The micro-and macrostructural design of improvised music," *Music Perception*, pp. 133-172, 1987.

[21] D. Conklin and I. H. Witten, "Multiple viewpoint systems for music prediction," *Journal of New Music Research*, vol. 24, no. 1, pp. 51-73, 1995.

[22] A. Smaill, G. Wiggins, and E. Miranda, "Music representation: Between the musician and the computer," in *Music education: an artificial intelligence approach*, M. Smith, A. Smaill, and G. Wiggins, Eds. Springer, 1993.

[23] M. Babbit, "Twelve-tone rhythmic structure and the electronic medium," Perspectives of New Music, vol. 1, no. 1, pp. 49-79, 1962.

[24] S. Eisenstein, *The Film Sense*. Harcourt, Brace and Co., 1947.

[25] A. Pickering, *The cybernetic brain: Sketches of another future*. University of Chicago Press, 2010.

[26] W. Grey Walter, *The living brain*. Gerald Duckworth and Co., Ltd., 1953.