Dory: a purposively flawed and forgetful artificial musical agent

Stefano Kalonaris

Queen's University Belfast
Sonic Arts Research Centre
skalonaris01@qub.ac.uk

Abstract. This paper presents Dory, a player-paradigm interactive music system for freely improvised music. While indebted to research in decentralised, agent-based musical systems and those employing Markov processes, Dory sets itself apart by capitalising on the artefacts introduced by suboptimal machine listening affordances and by speculating on the role of episodic and short-term memory in improvisation. Dory exhibits both reactive and learning traits, but is unable to exclusively commit to one or the other. By employing a subsumption architecture, unsophisticated machine listening techniques and probabilistic treatment of musical information, Dory appears as an engaging musical partner that displays seeming novelty and creativity, according to the author's preliminary qualitative evaluation.

Keywords: Computational Creativity, Markov Chain, Subsumption Architecture, Musical Free Improvisation, Episodic Memory, Short-term Memory.

1 Introduction

Interactive music systems in player-paradigm not under performer control and lacking enumerative rules are of particular relevance to this paper. The author's work is situated within the the context of such systems that have been developed to perform freely improvised music interactively with a human performer. Notable precedents include George Lewis' *Voyager* [1], Adam Linson's *Odessa* [2] and Tim Blackwell's *Swarm Music* [3]. These systems, although realised according to different methodologies, are fundamentally set apart from other player-paradigm systems such as corpus-based and those that try to emulate human performance and/or compositional practices through some degree of representation of human musical concepts. Amongst these, it is worth mentioning Robert Rowe's *Cypher* [4], François Pachet's *Continuator* [5] and Ircam's OMax system based on *factor oracle* [6]. One could use the above differentiation to divide player-paradigm systems into reactive and learning systems, depending on whether they are exhibiting emergent complex behaviour, or whether they account for corpus-based knowledge (acquired offline or in real-time) and/or driven by rule-based architectures. The first category would thus react to incoming audio stimuli, in dynamical and sometimes unpredictable ways, whereas the second would be able to mimic style or play in idiom. Dory is a player-paradigm interactive music system which tries to negotiate over the above methods and affordances, based on the author's desires and aesthetic beliefs/preferences. It is aimed for the performance of freely improvised music with a human co-player whose interaction experience and expert knowledge is paramount to the evaluation of the system itself.

Dory exhibits characteristics from both the above categories, and behaves according to two modes, which are not mutually exclusive. They are autonomously activated and operate in parallel, with signals between different layers of the system liable to be blocked or substituted. Before delving into the implementation details, a brief overview of the key concepts with regards to memory is presented.

2 Memory and Improvisation

Memory is often thought as divided into short and long term. Ranking under the second, one finds further layers, such as declarative (or explicit) and procedural (or implicit). Yet deeper into declarative memory, it is possible to discern between episodic and semantic memory. For the purpose of musical improvisation, studies have argued that procedural memory yields high speed (responsiveness) but low novelty, while declarative memory the contrary [7]. According to David Huron, episodic memory is linked to veridical expectations, which "represent invariant sequences learned from frequent exposure to a particular stimulus" [8, p. 363]. He associates short-term memory instead to dynamic expectation, contingent only on the contiguous experiences. In the context of an interactive music system for free improvisation, thus unbound from semantic and schematic representations of the music and which can alternatively account for speed and responsiveness, it seems appropriate to consider only short-term and episodic memory. In the process of representing these memory processes, the time bounds need re-scaling to appropriate values for a musical performance/experience. These details are discussed more in depth in the next section.

3 Proposed Model

Dory is entirely realised using Cycling74's Max programming language[1], built following the subsumption architecture popularised by Rodney Brooks [9] and, more specifically, inspired by Linson's *Odessa* [2]. Abiding by Brooks' approach, no attempt for accurate and semantic representation of the world, in this case the incoming audio, is made. Instead, a decentralised network of interaction between simple modules (akin to agents, in Marvin Minsky's terminology [10]) is preferred, which might induce emergent complex behaviour and privilege viewpoint decomposition [11], albeit without trying to make predictions. The conceptual link to the centrality of short-term and long-term memory in Conklin and

[1] https://cycling74.com

Witten's notion of viewpoint decomposition and the author's choices on memory described earlier is evident, but it is important to note that in the author's system the modelling of existing corpora through machine learning techniques is not contemplated nor desired. In terms of layer design, Dory comprises modules such as Listen, Play, Learn and Create. The last two are associated with two modi operandi, referred to as learning and creative modes. The inputs and outputs to/from the modules that implement these modes can be *inhibited* or *suppressed* depending on whether the signals are respectively blocked or replaced between layers/modules. The fundamental difference with Linson's system is that what is described as the adaptive agent in *Odessa*, is implemented in Dory with a partial learning goal. This might seem a small variation, but introduces an extra layer of interactivity, whereby Dory might elaborate over musical material it is acquiring from its collaborators and the environment around it. However, due to the considerable level of error in the representation of the live audio input (stemming from potential polyphonic input coupled with absence of source separation strategies and a purposeful underdevelopment of the 'listening' procedures), Dory's learning offers a rather partial impression of the real-time corpus. This is a crucial aspect that needs further explanation, both at a conceptual and technical level.

3.1 Listen

The first module/layer of Dory is where incoming audio is analysed and decomposed into three data streams, relating to pitch information, loudness and time deltas between notes detected. Furthermore, incoming audio's tempo is estimated, onsets detected and the chroma[2] set is calculated. Dory, unlike Rowe's *Cypher*, is not meant to receive MIDI information from the human player. The population of performers that the system targets, namely free improvisers, is often not very keen on MIDI controllers/instruments that might not only considerably change the players' symbiotic relationship with their instrument (e.g. hamper their ability, perceived or real, to be highly responsive and expressive) but also impact on their aesthetic beliefs on sound (often formed through years of dedicated practice). With these considerations in mind, Dory is meant to listen to acoustic or electro-acoustic instruments rather than MIDI. This, in turn, makes for the problematic task of estimating all the aforementioned parameters and features, particularly considering that free improvisation poses serious challenges to representation and classification goals. It is often part of the ideology of this musical expression to, in fact, resist formalisation. For a more in depth review of free improvisation please refer to Derek Bailey's writings [12].

3.2 Learn

When in learning mode (the corpus-based side of the system), Dory draws from two memory bins: one (very) short and the other as long as the performance.

[2] circular organisation of the pitch classes in the twelve-tone equal temperament

These two bins represent the earlier discussed short-term and episodic memory. As a result, Dory cannot learn off-line or learn extensively from her collaborators since she only occasionally draws from the longer memory pool. Dory can be said to lack implicit memory and she is corpus-based only within the scope of the performance itself. This trait is, in the author's opinion, crucial in forging parallels to how musical interaction in free improvisation might operate. In fact, it is neither within the scope of this musical expression to mimic or play in idiom/style, nor to establish an ex-novo vernacular within a performance. Although musicians might extrapolate some of the material they are presented with by the other improvisers, and elaborate on it to some extent, free improvisation thrives on the opportunities to "renew or change the known and so provoke an open-endedness which by definition is not possible in idiomatic improvisation" [12, p. 142]. The memory bins are implemented as four second order Markov chains, relating to pitch, loudness, time interval between notes detected from the incoming audio, and their duration, as processed in the Listen module. The short-term memory bin is initialised at 1 second; however, considering that human short-term memory is estimated in the range from a few seconds to a minute, this parameter can be experimented with and serve as an empirical value for Dory's sensitivity to memory. In learning mode, each of the sound parameters are obtained by querying the next value in their respective Markov chain, and they are then passed to the Play module, which might also receive data from the Create module.

3.3 Create

When Dory is in creative mode (the reactive side of the system), she disregards memory altogether and relies on probabilistic treatment of the chroma set and amplitude distributions. This mode could be also thought of as an impression, 'sense' or perception of the sonic input, which Dory elaborates on. The creative mode is normally activated when the level of event density of the incoming audio falls below a set threshold value. The weights (calculated as the incoming empirical distribution over the 12 pitch classes) of the chroma set are used to probabilistically trigger notes accordingly, while time intervals between the notes are obtained by randomly choosing standard subdivisions (semibreves, triplets, semiquavers, dotted notes, etc.) of the detected tempo from the incoming audio. The amplitudes of the notes are instead pooled from an arbitrary distribution.

3.4 Play

The Play module collects information sent by Learn and Create and produces the MIDI stream that controls the instrument chosen to voice Dory. The choice of having a sound agnostic output was motivated by the need for decoupling Dory from aesthetic judgement on the quality of the sound or sonic processes, which can become often obsolete and is in general too case-specific. When no audio activity is detected in the Listen module, Play receives a corresponding message that inhibits all of its inputs and that fades out Dory's output. Above a certain time interval of continued absence of stimuli, Dory shuts down.

4 Evaluation

Dory is currently being developed, and a formal evaluation procedure has yet to be designed. The evaluation of interactive music systems in player-paradigm presents considerable issues. This is even more true when dealing with such systems that are not under performer control. As a result, there seems to be a fundamental lack of vernacular in this respect. While the vast majority of developers and researchers focuses on quantitative methods, these seem to be inappropriate for validating systems aimed at, and performing with, free improvisers. Tempted by the Turing Test idea, Musical Directive Toy Tests (MDtTs), Musical Output Toy Tests (MOtTs) and Discrimination Tests (DTs) have been often used [13], ignoring the fact that these are oversimplifications of the Imitation Game [14] which, despite their often reported positive results, add very little to the artistic and aesthetic evaluation of the interactive music systems they target. In fact, these are at best an exercise in musical judgment rather than indicators and measures of thought, intelligence and creativity. Even less can these methods be considered informative in discerning aesthetic success. Particularly in the context of systems designed to produce freely improvised music,

> there are in general no clear goals, no criteria for testing correct answers, and no comprehensive set of well-defined methods. [15, p. 240]

Therefore, both the problem and the solution are ill-defined. As Lewis puts it,

> Avoiding scientism on the one hand and anthropomorphism on the other, I don't feel the need to 'scientifically' prove the validity of any process I use to get my music to sound the way I want it to sound. [16, p. 110]

Although often disregarded as not sufficiently scientific, qualitative evaluation might prove a strong contender as the primary methodology for validating player paradigm systems. In fact, expert opinion has been argued to be more appropriate for interactive music systems not under performer control and lacking enumerative rules.

> an experienced improvisor is well-suited to serve as an expert qualitative evaluator, capable of attunement to both subtle and complex emergent criteria. [17, p. 148]

It is also the author's opinion that expert qualitative evaluation is paramount and central to the discourse around interactive music systems that perform freely improvised music. Although still in prototypal stage and under current active development, Dory has shown (through performance with the author) promising characteristics and behaviours that suggest that she is an engaging musical improviser.

5 Conclusion

Dory was not motivated by the desire to model human musical thought nor semantic representations of musical structure or language, and it is therefore rather limited in this respect. The author's system was instead motivated by the desire to achieve *satisficing* [18] experiences of collaborative free improvisation with a human player by a system design implementation which would purposefully exploit the artefacts of simple machine listening strategies. To enhance the sense of interactivity with the human player, simple Markov processes were combined with a subsumption architecture, whereby a dynamical rumination/elaboration of the sonic world the system interacts with emerges. Without claiming to be a sophisticated agent, Dory is exhibiting encouraging dynamical traits that warrant further exploration in combining decentralised models with mnemonic affordances confounded by approximate musical feature extraction methods.

References

1. George E. Lewis. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, 10(2000):33–39, 2000.
2. Adam Linson, Chris Dobbyn, George E. Lewis, and Robin Laney. A subsumption agent for collaborative free improvisation. *Computer Music Journal*, 39(4):96–115, December 2015.
3. Tim Blackwell. Swarming and music. In *Evolutionary Computer Music*, pages 194–217. Springer London, 2007.
4. Robert Rowe. Machine Listening and Composing with Cypher. *Computer Music Journal*, 16(1):43–63, 1992.
5. François Pachet. The Continuator: Musical Interaction With Style. *Proceedings of the International Computer Music Conference*, 31(1):333–341, 2002.
6. Gérard Assayag, Georges Bloch, Marc Chemillier, Arshia Cont, and Shlomo Dubnov. OMax Brothers: A Dynamic Topology of Agents for Improvization Learning. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, AMCMM '06, pages 125–132, New York, NY, USA, 2006. ACM.
7. Christine Moorman and Anne S Miner. Organizational Improvisation and Organizational Memory. *The Academy of Management Review*, 23(4):698–723, 1988.
8. David Huron. *Sweet anticipation.* MIT Press, 2006.
9. Rodney A. Brooks. *Cambrian Intelligence: The Early History of the New AI.* A Bradford book. Bradford Book, 1999.
10. Marvin Minsky. *Society Of Mind.* Simon & Schuster, 1988.
11. Darrell Conklin and Ian H. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.
12. Derek Bailey. *Improvisation: Its Nature and Practice in Music.* Perseus Books Group, 1993.
13. Christopher Ariza. The interrogator as critic: The turing test and the evaluation of generative music systems. *Computer Music Journal*, 33(2):48–70, June 2009.
14. Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950.

15. Simon Holland. Artificial intelligence in music education: A critical review. In *Readings in Music and Artificial Intelligence*, pages 239–274. 2000.
16. George E. Lewis. Interacting with latter-day musical automata. *Contemporary Music Review*, 18(3):99–112, 1999.
17. Adam Linson, Chris Dobbyn, and Robin Laney. Critical issues in evaluating freely improvising interactive music systems. In *International Conference on Computational Creativity (ICCC) 2012*, pages 145–149, 2012.
18. Herbert Simon. Rational Choice and the Structure of the Environment. *Psychological Review*, 63:129–138, 1956.