Satisficing goals and methods in human-machine music improvisations: Experiments with *Dory*

Stefano Kalonaris

RIKEN Center for Advanced Intelligence Project, Tokyo

Abstract. Current interactive music systems for human-machine improvisation often employ sophisticated machine learning algorithms, achieving competence in style imitation and interaction with human performers within defined musical domains. However, in the context of free musical improvisation, it is probably not desirable to interact with a musical partner which one can largely anticipate or predict, because this might hinder the critical re-examination of one's improvisational practice, to the detriment of an open-endedness that is crucial in this particular form of musical expression. The author's contention is that, just as one would strive to collaborate with highly original and diverse musical personalities when freely improvising, a similar scenario would be desirable when collaborating with a computer system. By settling for "good enough" solutions to the problems posed by the design of the latter, and negotiating expectations of the attainable, a more unpredictable and contradictory agent might arise. In this article, the author's system's conceptual framework, design and an evaluation of three performances using it are presented.

Keywords: Computational Creativity, Human-machine Improvisation, Subsumption Architecture.

1 Introduction

The primary goal of the author is a system aimed at real-performance environments, where a free improvising musician interacts with a computational collaborator as a duet, during which the musician feels sufficiently engaged and stimulated in the musical interaction. While there exist many systems employing sophisticated methods such as artificial neural networks and deep learning architectures (Bretan et al., 2017; Eck & Schmidhuber, 2002; Hadjeres et al., 2017; Hutchings & McCormack, 2017; Mehri et al., 2016), the author considers here simple, unbaked and perhaps even naive solutions to musical problems so long as they meet his aspiration level, as he leans towards a *satisficing*[1] approach (Simon, 1956). The author aims to show that, by reducing the problem-solving scope of a system via specific design choices (e.g., methods for representing and generating musical input/output), an engaging musical agent might surface, appearing to

[1] A combination of "satisfying" and "sufficing".

negotiate between contradictory goals (i.e., reacting *versus* learning). While endorsing approaches that consider improvising computer systems decoupled from anthropomorphising desires and analogies (Bown, 2015; Lewis, 2000; Linson et al., 2015), the author's claim is that one would want to collaborate with someone/something that is sufficiently different yet sufficiently similar. Paradoxically, accepting "to go only part of the way towards satisfying a given value" (Cook, 1994, p. 7), which can translate to rough-hewn strategies and moderate search in problem-solving, might seemingly produce such a contradiction.

In this article, the author presents *Dory*, an interactive system for human-machine musical improvisation which is an example of employing satisficing methods and design, where "the search is for sufficient, not necessary, actions for attaining goals" (Simon, 1996, p. 124). After describing the motivations behind this endeavour, the context and the system design of *Dory*, the author evaluates this computational improviser with respect to three public performances. This suggest that *Dory* is best appreciated if considered in its own right as a computational collaborator, whereby it is no longer "technology or computers at all, but musicality itself" (Lewis, 1999, p. 110).

2 Frame

Computer systems that are able autonomously to interact musically with human performers are sometimes referred to as *live algorithms* (Blackwell et al., 2012), or interactive music systems in a *player-paradigm* (Rowe, 1992a). These autonomous musical agents are not devoid of the assumptions and beliefs of their makers, and are likely to reflect them in the way they are designed and operate. There is a rich legacy of music practitioners who have engaged with the task of designing, implementing and performing with such systems, and the one presented here is the author's contribution to this legacy. In particular, *Dory* is situated in the context of those systems (Hsu, 2008; Lewis, 2000; Linson et al., 2015) that are developed to perform freely improvised music which, according to some of its founders, it is meant to represent unbound *ex-novo* musical creation, devoid of formal structuralism and idiomatic rules. While claims of having no vernacular in terms of melodic, harmonic and rhythmic structures might be unrealistic, because every repeated artistic and aesthetic practice eventually leads to some formalisation, "much of the impetus towards free improvisation came from the questioning of musical language. Or more correctly, the questioning of the 'rules' governing musical language" (Bailey, 1993, p. 84). More specifically, the author is interested in those systems that are not under player control, meaning that the human performer has no access to the internal parameters during performance.

Since the musical domain for which *Dory* is intended has no formal rules in terms of structures or idiom, the author's design was freed from considerations regarding high-level representations of music. These could include embedding musical knowledge or rules regarding melody, harmony and rhythm, as well as formal structures such as bars, musical phrasing, periods, hierarchical relationships and thematic development, for example. Such representations would be paramount for other systems aimed at style imitation, which would normally be trained extensively over large corpora, leading to a peculiar scenario where the machine might well have "listened" to more music than its human collaborator. Furthermore, those systems often consider a wide range of options before committing to a musical decision, perhaps solving complex optimisation and inferential problems, and finding the best possible path or sequence given the data available. Not only are these unrealistic characteristics for a human improviser, who is too slow to do exhaustive systematic search, but it is perhaps even undesirable to collaborate with a musical agent which is infinitely competent, has memory of an arbitrarily large corpus, and is able to actuate the best response to a given problem. In other words, "it might be intrinsically rational not to seek, and to be satisfied with some 'aspiration level' of results less than the achievable best" (Slote, 2004, p. 28). Settling for coarse and rough-hewn strategies might, for example, help draw analogies with non-rational characteristics of human decision-making.

Naturalistic decision-making (NDM) studies (Klein et al., 1993) have shown that, under time and pressure constraints, humans are likely to opt for a "good enough" solution. This is akin to the concept of *bounded rationality*, first introduced by Herbert Simon (1956), who also coined the term satisficing. These ideas posit that, providing that one can define a minimum threshold with respect to given needs, the first solution that meets such criterion is that chosen. This threshold has been referred to as *aspiration level*, comparing the "expectations of the attainable ... with the current level of achievement" (Simon, 1996, p. 30). Speculatively borrowing from this theory, the author's aspiration levels were defined by several factors which are thought to be crucial in determining the appreciation and/or success of human-machine musical interaction. According to widespread opinions (Brown et al., 2017; Hsu & Sosnick, 2009; Peters et al., 2009), a system in this context should be: *responsive*, *stimulating*, *engaging* and *surprising*. While the first two factors will be considered exclusively in the context of the performer's perspective (see Section 4.3), the last two were also used as measures of the audience response (see Section 4.2).

A satisficing approach was also applied to the methods and the design of *Dory*. In particular, acceptable rather than best possible solutions were adopted with respect to the system's faculty of listening and learning. By these terms the author specifically refers to *machine listening* and *machine learning*, respectively. Concretely, *Dory* is designed to listen and process incoming raw audio instead of MIDI information, while completely disregarding source separation and polyphonic input issues. A more detailed discussion is deferred to Section 3.2. As a counter-example, one could imagine systems such as *Cypher* (Rowe, 1992b), which has an extremely accomplished *listener* module. This is achieved partly due to the fact that the incoming information is provided in MIDI format, thus presenting little, if any, room for ambiguity. Furthermore, *Dory* is unable to learn extensively or train from corpora or live input, due to its short memory affordance. This deliberate design feature arose from speculations on how short-term memory reflects on improvisation (Moorman & Miner, 1988), and it is explained more in depth in Section 3.3. A contrasting example of an accomplished *learner* would be Pachet's *Continuator* (2003). *Dory* is thus liable to an arbitrarily fuzzy recognition of the "world" and musical misinterpretations of the human co-improviser(s). While these properties might seem at odds with a coherent musical output, writer and musician Rob Wallace reminds us that "a more nuanced view of improvisation reveals ... a constantly negotiated framework of sometimes contradictory possibilities" (2010, p. 34). To understand better how *Dory*'s behaviour emerges from the symbiosis of contradictory possibilities, it is crucial firstly to account for some important design considerations.

3 System Design

Although a thorough discussion regarding common design strategies is beyond the scope of this article, it is important to identify some essential design factors that can provide a compact description of computational musical collaborators. This can be particularly useful for comparing different systems or different versions of the same system, for example. The factors that can more succinctly illustrate what specific choices were made in the design of *Dory* are the *architecture*, the type of *input/output*, the *model type*, and the *methods* used to generate musical material.

Architecture relates to how the system is built. Examples could be multi-agent architectures that simulate swarm intelligence (Blackwell, 2007) or complex interaction of independent music bots (Eigenfeldt et al., 2015). Input/output simply describes what data type the system is reading/processing and generating. The model type specifies whether the system is able to learn from a corpus, live input or both (Assayag et al., 2006; Hadjeres et al., 2017; Pachet, 2003; Young, 2008), whether it draws on embedded musical knowledge (Rowe, 1992b), or whether it is instead reactive (Bown et al., 2014; Lewis, 2000) or reflexive (Pachet et al., 2013; Weinberg et al., 2008). In other words, the model type is closely related to the motivations of the system designer. Methods used to generate musical output are diverse, including grammar-based (Gillick et al., 2010), L-systems (McCormack, 2003), knowledge-based and self-organising maps (SOM) (Rowe, 1992b), artificial neural networks (Manzolli & Verschure, 2005), evolutionary algorithms (Biles, 2002; Weinberg et al., 2008), and Markov processes (Wang et al., 2016). For a comprehensive review of algorithmic methods, the reader is encouraged to consult recent surveys (Fernández & Vico, 2013; Herremans et al., 2017). Armed with these four definitions, it is now possible to describe *Dory* accordingly.

3.1 Architecture

Dory is built following a distributive and parallel control design, whereby different modules can be activated concurrently, thus negotiating different levels of agency. It is rooted in Rodney Brooks' (1999) stance, which stands in opposition to emulating symbolic reasoning and mental representations. This challenged the up-to-then dominant model of the *sense-plan-act* (SPA) strategy for implementing

AI systems by entirely removing the *plan* stage. In the early 1990s Brooks argued for a sensory-motor coupling with the environment, connecting limited task-specific perception directly to required actions, and he built several robots following this radical approach, employing a decentralised network of interaction between simple modules (akin to agents, in Marvin Minsky's terminology (1986)). This architecture is referred to as *subsumption* and inspired a wealth of later experiments (Hasslacher & Tilden, 1995; Quigley et al., 2009). Subsumption is based on a few key concepts, such as *situatedness* (whereby the world is its own best model), *embodiment* (having a concrete, integrated physical control system, with direct coupling of sensor data to actions), *intelligence* as a by-product of bottom-up approaches (mobility, locomotion and perception are the first foundations of intelligence), and *emergence* (systemic behaviours arise from the interaction of smaller local entities). In such systems the inputs and outputs to/from the modules can be *inhibited* or *suppressed* depending on whether the signals are, respectively, blocked or replaced between layers. These modules operate in parallel and guarantee that the system will always function at some level. It is understood that higher levels *subsume* lower ones. For example, in Brooks' robots the lowest operational level would be "avoid obstacles" and a higher one could be "move towards target".

Studies related to dynamical systems and embodied cognition (Chemero, 2009; Clark, 1996), some of which have parallels in music (Borgo, 2005; Di Scipio, 2003), might exhibit similarities to Brooks' approach, due to the connectionist model underlying them and to the shifted focus towards distributed, multidimensional, local and action-oriented representation, as opposed to being objective and action-independent. In terms of notable precedents in employing a subsumption architecture in the context of human-machine musical interaction, the work of Linson (Linson et al., 2015) stands out. When translating Brooks' concepts to a musical system, bespoke choices were made for *Dory.* The author's system comprises 7 modules, with *listen* being the lowest level of operation (discussed in Section 3.2). A schema of *Dory*'s architecture can be seen in Fig. 1.

Two of these modules (*follow* and *remember*) explicate memory functions, while two others (*feel* and *create*) are responsible for more reactive traits of the system (these will be discussed in Section 3.3). A more in-depth flow diagram of *Dory*'s architecture is shown in Fig. 2. The next system characteristic to be described relates to how it receives and outputs signals.

3.2 I/O

For accurate modelling of incoming musical information, many systems tend to be designed to accept input as MIDI data. Furthermore, some impose further restrictions in this sense, and are optimised to deal with monophonic instruments (Biles, 2002). These strategies ensure that segmentation and representation can eventually reflect the input to a high degree of accuracy. *Dory*, by contrast, has no such strategies. Not only is it unconcerned with polyphony, but it is also designed to accept direct audio from the musician's instrument or the environment. Such a deliberate approach produces a considerable level of error in the representation

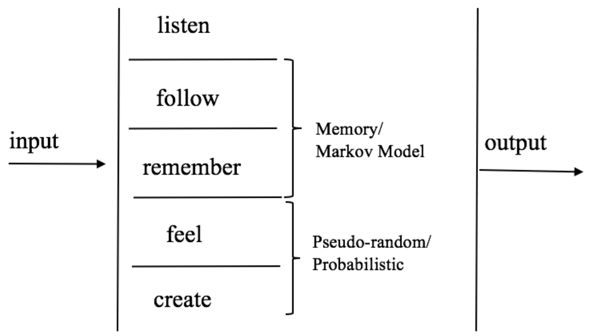

Figure 1. A subsumption architecture. Modules/layers are activated in parallel and each connects the input to the output. Adapted from Brooks (1999, p. 67).

of the live audio input, further compromising *Dory*'s already partial learning capabilities (discussed in Section 3.3).

In the *input* module, raw audio from the human player is received and a simple averaging of the peak amplitude over three time windows (2500ms, 5000ms and 7500ms) with respect to a *sensitivity* hyper-parameter is computed, as a list of three binary values (e.g., 0 0 1). This list is a measure of how "busy" the human player is, sonically. If no activity is detected within 7500ms then *Dory* hibernates and the *output* module is inhibited. The binary lists so obtained are used to determine which of two finite state machines (FSMs) should be queried. These FSMs are in charge of deciding the activation state for *create/follow* and *feel/remember*, respectively (discussed in Section 3.3). The transition weights can be adjusted, to confer a more or less reactive or learning personality to *Dory*. However, this can only be done offline, and once the system is operational, the human performer no longer has control over these weights.

The audio input is then fed to the *listen* module, where it is analysed and decomposed into three data streams, relating to pitch information, amplitude and time deltas between notes detected. This is done by estimating the fundamental frequency of the incoming audio by performing multiple layers of wavelet transform on an incoming vector, comparing the spacing between the peak in each. The peak amplitude is converted to a range of 0–127, used as MIDI velocity values. Similarly, pitch values are also converted to MIDI notes. As for the time deltas between the notes, these are calculated simply as the time elapsed between two consecutively detected onsets. The information so obtained will populate the corresponding transition matrices of the Markov chains in the *follow* and *remember* modules (see Section 3.3). The object that performs this analysis (part of the standard distribution of the software[2] used to implement *Dory*) is specifically optimised for monophonic signals. This specification is, however,

[2] *Max*; at https://cycling74.com.

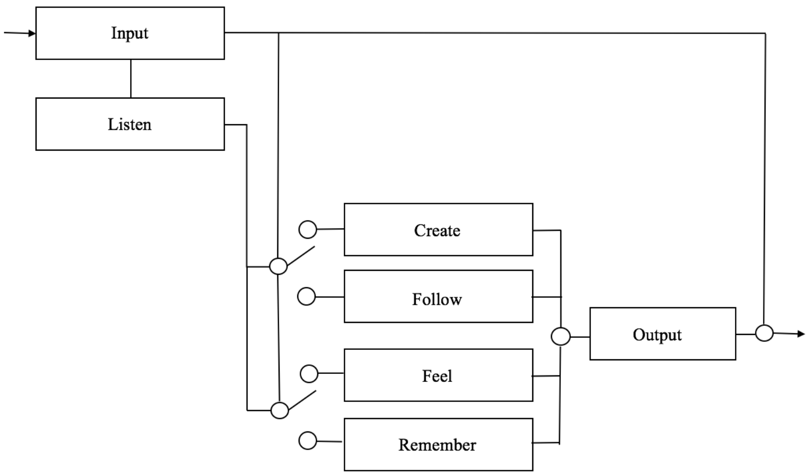

Figure 2. *Dory*'s flow diagram. Modules are represented by named boxes and solid lines represent the data flow. Clearly visible are the two FSMs discussed in Section 3.2, determining which module(s) will be activated/inhibited.

purposively disregarded when *Dory* is playing with a human using a polyphonic instrument, resulting in an arbitrarily approximate and incorrect representation of the incoming audio. Additionally, the incoming audio's tempo is estimated via a real-time beat-tracking model (Stark et al., 2009), and the chroma set[3] is calculated (Stark & Plumbley, 2009). This information is then used in the *feel* and *create* units (see Section 3.3).

The type of output that *Dory* produces is in MIDI format which can be sent to an external instrument (hardware or software) of the user's choice. Thus, *Dory* does not have the means to generate actual sounds, but only data streams. This strategy is common amongst system designers, because it frees them from choices relating to sounds which could become aesthetically obsolete or historically bound. In doing so, a much wider palette of sounds is made available to the human improviser, who might choose different "players" for different performances or occasions. Once again, this choice can only be made before the system is activated, after which point it is no longer possible to change *Dory*'s output sound. The next characteristics to be considered will be the system's model type and the methods used to generate MIDI output.

[3] The circular organisation of the pitch classes in twelve-tone equal temperament.

3.3 Model Type and Methods

Dory can be said to be both reactive and learning, albeit unable to commit to one or the other. This is a result of the system's architecture, as seen in Section 3.1.

The reactive aspect of *Dory*'s behaviour is obtained via the activation of the two modules named *create* and *feel*, which generate streams relative to pitch, time deltas between notes, amplitude and duration of the notes. This process is largely based on statistical distribution sampling, coupled with rhythmic subdivision based on the detected tempo. These two modules are similar, except that *feel* computes an arbitrarily long averaging over incoming note speeds, chroma content and detected tempo. The time window over which these are calculated is the time between the last and the current activation of the module itself. The *create* module, by contrast, does no averaging and simply computes the same elements as a running stream. In both modules, the duration and amplitude of the notes are pooled from an arbitrary distribution.

The learning side of the system is realised via Markov processes and is also implemented according to two different time scales, reflecting considerations made in an earlier paper by the author (Kalonaris, 2017), where a model of memory theorised by Tulving (1972) was considered (shown in Fig. 3). Following studies that have shown a positive correlation between declarative memory and high novelty (but low speed) in music improvisation (Moorman & Miner, 1988), a case was then made for implementing *Dory* with a partial learning goal, using only two short bins of memory which represented the workings of short-term and episodic memory in the Tulving model. This choice was further supported by the findings of Huron (2006) and Snyder (2000).

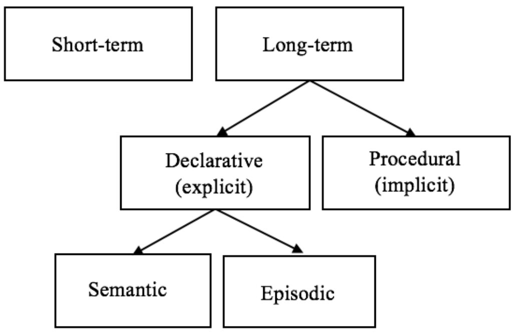

Figure 3. Tulving's (1972) memory model, hypothesising two kinds of long-term memory: declarative and procedural. While the latter is associated with implicit and subconscious recall, the former is responsible for conscious and intentional recall, and it is thought to be further divided into semantic and episodic memory.

Data regarding pitch, amplitude and time delta received from the *listen* module (described in Section 3.2) is used to populate three corresponding transition matrices, in either a short memory module (*follow*) or a longer one (*remember*, representing episodic memory). The former is wiped every time the module is not active, during which period the transition matrix is filled with the new incoming streams. The latter was originally implemented to run for the duration of the performance, but its length can be heuristically reduced by the human performer, similarly to the sensitivity hyper-parameter (see Section 3.2) . When either module is active (i.e., when their output is not inhibited) second-order Markov chains are queried for the next values to be sent to the *output* module.

As a result of this limited memory affordance, *Dory* is incapable of learning over a real-time corpus, which explains the system's inability to model the human collaborator's musical language. However, some local and time-sensitive characteristics might at times be reflected in the output of the system. In other words, *Dory* might occasionally and briefly elaborate on musical material derived from its human collaborator, or even from the environment around it (if audio is acquired via a microphone). Adding to the reduced ability to learn extensively is the fact that what the system feeds to the two memory bins might already be spurious and compromised to some degree, due to the flaws of the *listen* module (see Section 3.2). Having described the anatomy and design philosophy of *Dory*, the preliminary results of the author's experiments in real-performance environments with the system are now presented.

4 Evaluation

Evaluation of computationally creative music systems poses many problems and it is an open discourse voicing many perspectives, trends and approaches. In the context of autonomous musical agents such as those described in this article, one can identify a "first wave" of such systems which were eminently practice-based and whose designers were not perhaps particularly concerned with formal evaluation methodologies. A later stage saw attempts to evaluate these systems with tests inspired by the Imitation Game (Turing, 1950). These comprised Musical Directive Toy Tests (MDtTs), Musical Output Toy Tests (MOtTs) and Discrimination Tests (DTs), for example, although there seems to be a consensus that such tests seem to fall short of being able to say anything about the intelligence or creativity of the systems (Ariza, 2009) or about the aesthetic and artistic results. At best, they end up being measures of musical judgement. At times the need for evaluation has been challenged altogether, because "what purpose would be satisfied by creating qualitative criteria or quantitative metrics for artificial musical intelligence, given the lack of successful similar criteria for natural musical intelligence, musicality, or even music per se?" (Belgum et al., 1988, p. 9).

It would, however, be undesirable to lack completely the possibility to compare systems, thus being able incrementally to improve knowledge in the domain and build on the existing legacy. While some practitioners seem to be leaning towards qualitative methods (Banerji, 2012; Bown, 2015), others are more inclined towards a mixture comprising both quantitative and qualitative measures (Stowell et al., 2009), with Human Computer Interaction (HCI) criteria also being common in this context (Hsu & Sosnick, 2009). Not only have attempts been made to design standardised procedures for comparing systems (Colton, Charnley, & Pease, 2011; Jordanous, 2012), but some also argue for the formal evaluation of different versions of the same system (Colton et al., 2014).

Dory is currently being developed, and a formal evaluation procedure has yet to be designed. The author shares many of the preoccupations presented above, and although he has not yet designed a rigorous evaluation method, he has attempted some preliminary evaluation of three public performances with *Dory*. It is important to note that he argues for systems designed for real-performance environment with experienced practitioners, much in accordance with other researchers' views (Bown, 2015; Lewis, 2000; Linson et al., 2015; Hsu & Sosnick, 2009).

4.1 Methodology

Dory was used by the author in three performances, which took place at the *Improvisational Creativity Workshop*, Prato, July 20, 2017; at the *2nd Conference on Computer Simulation of Musical Creativity*, Milton Keynes, September 13, 2017; and at the *New River Studios*, London, October 25, 2017. These were all scenarios where a rigorous, formal evaluation may have not been appropriate (concert environment, low response rate, etc.), however, both a quantitative study of the audience response and a descriptive analysis from the performer's point of view are provided. The former relates to the first two performances, whereas the latter is cumulative over all three.

4.2 The Audience's Perspective

Evaluation methods for creative systems have been divided into methods of external and internal evaluation (Agres et al., 2016). Amongst the former, both questionnaires and correlation tests feature as valid options. In particular, it is argued that Likert rating scales, although subjective, "can provide very robust and consistent measures" (Agres et al., 2016, p. 17). Given the context (a real-performance environment) and the above consideration, a simple questionnaire was used. This consisted of only three items:

- I enjoyed the performance (measure of *enjoyment*)
- The musical interaction was sufficiently engaging (measure of *engagement*)
- *Dory* sometimes responded in unexpected ways (measure of *surprise*)

A ten-point Likert scale ranging from "strongly disagree" to "strongly agree" (re-coded 1–10) was used to quantify the audience response. As already noted, engagement and surprise are, within the computational creativity community, consensually agreed upon as important factors in the evaluation/determination of a system's creative ability. Therefore, these two measures were used for the second and third questionnaire items, respectively, whereas the first item is asking whether both the interaction between the human performer and the system, and the resulting musical output, were considered aesthetically effective and/or pleasing. Audience members were all domain experts and they were participants of the two events (July 20 and September 13, 2017), either presenting papers on creative computational systems or performing as part of the concerts. Amongst them, some of the most eminent practitioners of human-machine improvisation were present. One of them provided written feedback in a private email. In total, only 9 participants responded.

Mean and median values for all three items were very high, as seen in Tab. 1. Despite a strong positive correlation between enjoyment and engagement, a chi square test of independence showed that the relation between these two variables was not significant.

	Enjoyment	Engagement	Surprise
Min	7	5	7
1st Qu.	8	8	9
Median	8	8	9
Mean	8.3	7.7	9.0
3rd Qu.	9	8	10
Max	10	10	10

	Enjoyment	Engagement	Surprise
Enjoyment	1.0	0.77	-0.17
Engagement	0.77	1.0	0.0
Surprise	-0.17	0.0	1.0

Table 1. Basic descriptive statistics (left) and correlation matrix (right) for N=9 participants in 2 different performances, answering three questions using a 1–10 Likert scale.

Some of the respondents volunteered comments and suggestions, some of which indicated a high level of surprise (reflected in Tab. 1):

> It surprised me in many ways – perhaps some graphical interaction with audience could help as *Dory* makes (her/his/its) decisions.

At other times a sense of wonder, considered "intimately connected with creativity" (Boden, 2003, p. 277), was expressed:

> It was difficult for me to get into your abstract communication with the program – and that is maybe a good sign that things are not too obvious.

Some comments also highlighted the short-term feedback of musical material between the author and the system:

> I found a couple of central notes from you and *Dory*, that kept coming back and heard you taking over ideas from *Dory* – my favourite moments.

Given the small number of respondents and the lack of a robust survey design method, the above results should be interpreted with caution and complemented with a performer study.

4.3 The Performer's Perspective

The author is a guitarist and electronic musician experienced in free improvised music. It has been claimed that "expert qualitative analysis should be recognised as fulfilling an essential role" (Linson et al., 2012, p. 148) in evaluating musical improvisations, therefore, despite the bias of the author being both the designer and the performer, a descriptive study is now offered. The standpoint of the following reflection is that of a practice-based approach, in full awareness that it might make more sense to consider it as instrumental to future implementations of the system. *Dory* was designed to achieve stimulating and engaging interactions with the human by freely improvising musical duets. Studies on creative partnership in human-machine musical improvisation (Brown et al., 2013, 2017) have identified some essential activities and relationships in a duet interaction. These are:

- *initiate* (new material)
- *imitate* (reuse of other's material)
- *loop* (immediate reuse of own material)
- *restate* (reuse of other's material over a longer time span)
- *shadow* (play in unison or close parallel)
- *silence* (not playing)

Such activities are intrinsically bound to the factors identified in Section 2 (i.e., that a system should be *responsive*, *stimulating*, *engaging* and *surprising*), and are discussed in relation to these. For example, initiating new or contrasting material, as well as not playing at times, is related to how the performer perceives the system in terms of being stimulating, engaging and surprising, while the capacity to imitate might be more associated with the responsiveness of the system.

Dory's ability to *initiate* musical material, at times in discordance with the human performer's musical behaviour, can be seen in Fig. 4. In this case, *Dory* starts with a new figuration in (perceived) quavers and a strong tonal centre of D minor, following a few seconds of silence observed by both the system and the author. Despite the variety of the author's response (firstly doubling the rhythmic figurations, then introducing bursts of four-note chordal support for *Dory*'s lines), the system did not stray from its musical direction. This characteristic is often much appreciated and sought-after when improvising, particularly in a duet. Another way to interpret it is as the ability to stray from reflexive interaction by means of either 'stick to your own' or introducing a contrastive musical idea. An example of the latter behaviour is illustrated in Fig. 5, where, after a short interchange of syncopated phrasing, *Dory* picked the perceived tonic and held it for a few seconds. Naturally, this provided an ideal opportunity for the author to change the texture of his musical ideas, as the music seemed quickly to transition to a low-energy state, allowing for more meditative ideas to emerge. At other times, similar behaviours of *Dory* elicited instead a different reaction in the author, whereby he gained the foreground and used fast divisions and/or complex ideas on top of *Dory*'s static and persistent counterpart.

Figure 4. An example of *Dory*'s ability to initiate a musical idea independently from the material played by the author.

These musical behaviours make for engaging interactions and the author, after an initial period of familiarisation, was able to negotiate his expectations relating to the system and accept *Dory* as an active element of the musical experience. While relinquishing ideas of humanised behaviours, the author did not perceive *Dory* simply as a "prop" or as a device/strategy for augmenting the sonic palette of the performance but, rather, as a fully fledged agent which contributed to it on an equal footing. He regarded the system as stimulating and engaging, because it helped him focus on musical interactions at a purely sonic level (given that *Dory* lacks a physical dimension). This perception was shared by some of the audience members, as shown in the following remark:

> I had to focus more on the listening as there was no performer to look at.

Due to the idiosyncratic behaviours of the system, the author was prompted to examine some of his preconceptions about interactivity in music improvisation and to adopt an attentive listening approach. According to some influential practitioners (Corbett, 2016; Solomon, 1986), rote imitation and polarisation can be problematic in free improvisation, resulting in predictable, concerted patterns of arousal and decay of musical energy. *Dory*'s lack of pure imitation was not a result of accomplished musicianship and years of experience in free musical improvisation, but was rather hardcoded in by virtue of the already discussed design features. Nevertheless, the system was able to exhibit short-range imitative traits that at times resulted in subtle interaction showing both harmonic and mood integrity, as can be seen in Fig. 6. On this occasion, during the third performance of the author with the system, the music maintained a strong pivot of C mixolydian. *Dory*'s long notes provided the author with the

Figure 5. An example of *Dory*'s ability to produce and maintain diverging musical ideas.

opportunity for a harmonic commentary, where he adopted a supportive role, as well as trading short melodic interventions at times. The author was able to experiment with dynamics, which allowed him to perform solo for brief moments. Such an affordance is contingent upon *Dory*'s sensitivity hyper-parameter, which sets an amplitude threshold for incoming audio below which even the *listen* module is suppressed (see Section 3.2). This particular behaviour is referred to as *silence* in the crucial factors of a duet interaction listed at the beginning of this section. The sonic texture need not be congested at all times and, in fact, the author appreciated the opportunity to "prune" the density of the music at times. Although this resulted from a design feature of the system, it was still felt and perceived in terms of a positive musical affordance, in the context of the performances. This behaviour is illustrated in Fig. 7.

5 Discussion

Despite the many positive qualities of *Dory*'s behaviour, the author also experienced some evident shortcomings. For example, the ability to produce entrainment, musical synergy and to deliberately sustain musical interaction is also crucial in a creative duet, since novelty and heterogeneity might not be sufficient to engage a human performer musically and aesthetically over a longer time span. In this sense, the system did not exhibit such capability. This lack of long-term structure and entrainment has also been noticed by improvisers in relation to other similar systems (Bown, 2015) and suggests that more can be done in the context of reactive systems with limited or no learning capabilities. Other shortcoming of the author's system were the lack of three of the activities listed at the beginning

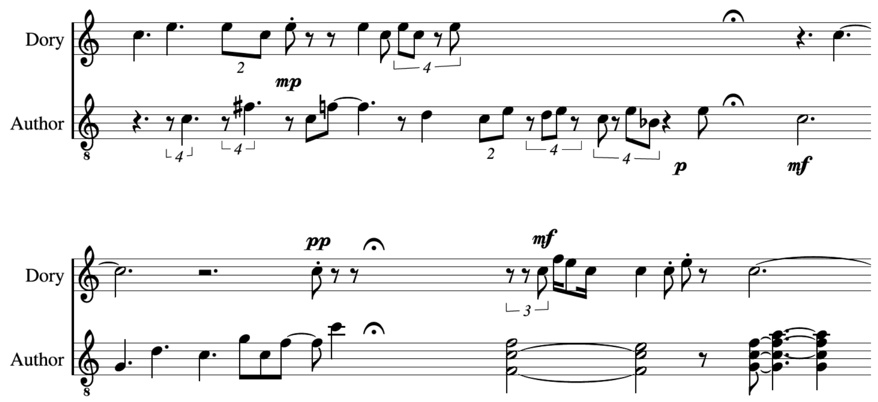

Figure 6. An example of musical imitation between *Dory* and the author, with a C mixolydian modal anchor.

Figure 7. An example of sparse interaction and use of silence on the part of *Dory*.

of Section 4.3, namely *loop*, *restate* and *shadow*. To address these issues, several strategies could be employed. Regarding entrainment and the immediate reuse of its own material, the system could, for example, be augmented with closed-loop (feedback) methods which could be triggered with respect to the level of metric or chroma consistency detected in the human player. However, since rote imitation and parroting can be equally undesirable in music improvisation (see Section 4.3), such affordance should be carefully implemented. Perhaps this could be done by including some mechanism for simulating creative divergence, along the lines of that discussed already (i.e, contrastive ideas; see Fig. 5), in order to escape loops seemingly by the system's initiative. To achieve the ability to *restate* it would be sufficient to stretch the memory length of the *remember* module and/or to alter the weights of the FSMs (see Section 3.2) to make the transitions to the long-term memory more probable. Another strategy would be to use more sophisticated algorithms for the Markov processes, for example Factor Oracle (Assayag & Dubnov, 2004). Such alterations would, however, be in stark contrast to the initial motivations behind *Dory* and, therefore, inappropriate for this particular system.

6 Conclusion

Dory was developed to be used for pairwise musical interaction and for freely improvised music, in real-performance environments. While it is tempting to desire human-like attributes in computational systems for music improvisation, it is important to negotiate these expectations with the awareness that these idiosyncratic systems might be better employed when they are granted bespoke status. By maintaining an open approach to performing with these systems, the human performer might not only learn dynamical couplings with the system, but might also be presented with opportunities for a critical re-examination of his/her improvisational and creative practices.

Dory set out to represent an example of satisficing design and of purposively reduced scope on several fronts. These choices were made following specific considerations regarding memory, perception and decision-making, and were realised via coarse machine listening and learning modules, as well as naturally emerging from the system's architecture. The latter was such that *Dory* lacked a concept of "planning", relying instead on distributed and parallel control and agency. Despite the limitation of the current implementation, such as the inability to sustain long-term musical interaction, *Dory* was received positively by expert audiences and by the human player alike, and it was deemed an engaging musical agent, exhibiting novel and idiosyncratic behaviours. These encouraging preliminary results warrant further exploration in this direction and the author intends to develop more formal methodologies for evaluating the system as well as strategies for attaining long-term capabilities that could perhaps simulate the desire to adapt to shared musical goals.

References

Agres, K., Forth, J., & Wiggins, G. A. (2016). Evaluation of musical creativity and musical metacreation systems. *Computers in Entertainment*, *14*(3), 3:1–3:33. doi: 10.1145/2967506

Ariza, C. (2009). The interrogator as critic: The Turing test and the evaluation of generative music systems. *Computer Music Journal*, *33*(2), 48–70. doi: 10.1162/comj.2009.33.2.48

Assayag, G., Bloch, G., Chemillier, M., Cont, A., & Dubnov, S. (2006). OMax brothers: A dynamic topology of agents for improvization learning. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia* (pp. 125–132). New York, NY: ACM. doi: 10.1145/1178723.1178742

Assayag, G., & Dubnov, S. (2004). Using factor oracles for machine improvisation. *Soft Computing*, *8*(9), 604–610. doi: 10.1007/s00500-004-0385-4

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

Banerji, R. (2012). Maxine's Turing test: A player-program as co-ethnographer of socio-aesthetic interaction in improvised music. In M. Riedl & G. Sukthankar (Eds.), *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment International Conference.* Palo Alto, CA: AAAI Press. Retrieved from `https://www.aaai.org/ocs/index.php/AIIDE/AIIDE12/paper/download/5533/5739`

Belgum, E., Roads, C., Chadabe, J., Tobenfeld, T. E., & Spiegel, L. (1988). A Turing test for "musical intelligence"? *Computer Music Journal*, *12*(4), 7–9.

Biles, J. A. (2002). GenJam: Evolution of a jazz improviser. In P. J. Bentley & D. W. Corne (Eds.), *Creative evolutionary systems* (pp. 165–187). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Blackwell, T. (2007). Swarming and music. In E. R. Miranda & J. A. Biles (Eds.), *Evolutionary computer music* (pp. 194–217). London: Springer. doi: 10.1007/978-1-84628-600-1_9

Blackwell, T., Bown, O., & Young, M. (2012). Live algorithms: Towards autonomous computer improvisers. In J. McCormack & M. d'Inverno (Eds.), *Computers and creativity* (pp. 147–174). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-31727-9_6

Boden, M. A. (2003). *The creative mind: Myths and mechanisms* (2nd ed.). New York, NY: Routledge.

Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age.* New York, NY: Continuum.

Bown, O. (2015). Player responses to a live algorithm: Conceptualising computational creativity without recourse to human comparisons? In H. Toivonen, S. Colton, M. Cook, & D. Ventura (Eds.), *Proceedings of the International Conference on Computational Creativity* (pp. 126–133). Park City, UT: Brigham Young University.

Bown, O., Gemeinboeck, P., & Saunders, R. (2014). The machine as autonomous performer. In L. Candy & S. Ferguson (Eds.), *Interactive experience in the digital age: Evaluating new art practice* (pp. 75–90). Cham: Springer International Publishing. doi: 10.1007/978-3-319-04510-8_6

Bretan, M., Oore, S., Engel, J., Eck, D., & Heck, L. (2017). Deep music: Towards musical dialogue. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence* (pp. 5081–5082). San Francisco, CA: AAAI Press.

Brooks, R. A. (1999). *Cambrian intelligence: The early history of the new AI.*

Cambridge, MA: MIT Press.

Brown, A. R., Gifford, T., & Voltz, B. (2013). Factors affecting audience perceptions of agency in human computer musical partnerships. In E. Y. Do, S. Dow, J. Ox, S. Smith, K. Nishimoto, & C. T. Tan (Eds.), *Proceedings of the 9th ACM Conference on Creativity & Cognition* (pp. 296–299). New York, NY: ACM. doi: 10.1145/2466627.2466661

Brown, A. R., Gifford, T., & Voltz, B. (2017). Stimulating creative partnerships in human-agent musical interaction. *Computers in Entertainment*, *14*(2), 5:1–5:17. doi: 10.1145/2991146

Chemero, A. (2009). *Radical embodied cognitive science.* Cambridge, MA: MIT Press.

Clark, A. (1996). *Being there: Putting brain, body, and world together again* (1st ed.). Cambridge, MA: MIT Press.

Colton, S., Charnley, J., & Pease, A. (2011). Computational creativity theory: The FACE and IDEA descriptive models. In D. Ventura, P. Gervás, D. F. Harell, M. L. Maher, A. Pease, & G. Wiggins (Eds.), *Proceedings of the 2nd International Conference on Computational Creativity* (pp. 90–95). México City. Retrieved from `http://computationalcreativity.net/iccc2011/proceedings/the_foundational/colton_1_iccc11.pdf`

Colton, S., Pease, A., Corneli, J., & Cook, M. (2014). Assessing progress in building autonomously creative systems. In *International Conference on Computational Creativity* (pp. 137–145).

Cook, J. (1994). Agent reflection in an intelligent learning environment architecture for musical composition. In M. Smith, A. Smaill, & G. A. Wiggins (Eds.), *Music education: An artificial intelligence approach* (pp. 3–23). London: Springer.

Corbett, J. (2016). *A listener's guide to free improvisation.* Chicago, IL: University of Chicago Press.

Di Scipio, A. (2003). 'Sound is the interface': From interactive to ecosystemic signal processing. *Organised Sound*, *8*(3), 269–277. doi: 10.1017/S1355771803000244

Eck, D., & Schmidhuber, J. (2002). Finding temporal structure in music: Blues improvisation with LSTM recurrent networks. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing* (pp. 747–756). Martigny, Switzerland: IEEE. doi: 10.1109/NNSP.2002.1030094

Eigenfeldt, A., Bown, O., & Casey, B. (2015). Collaborative composition with creative systems: Reflections on the first musebot ensemble. In H. Toivonen, S. Colton, M. Cook, & D. Ventura (Eds.), *Proceedings of the International Conference on Computational Creativity* (pp. 134–141). Park City, UT: Brigham Young University.

Fernández, J. D., & Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research*, *48*(1), 513–582.

Gillick, J., Tang, K., & Keller, R. M. (2010). Machine learning of jazz grammars. *Computer Music Journal*, *34*(3), 56–66. doi: 10.1162/COMJ_a_00006

Hadjeres, G., Pachet, F., & Nielsen, F. (2017). DeepBach: a steerable model for Bach chorales generation. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (pp. 1362–1371). Sydney, Australia: PMLR.

Hasslacher, B., & Tilden, M. W. (1995). Living machines. *Robotics and Autonomous Systems*, *15*(1–2), 143–169.

Herremans, D., Chuan, C., & Chew, E. (2017). A functional taxonomy of music generation systems. *ACM Compututing Surveys*, *50*(5), 69:1–69:30. doi: 10.1145/3108242

Hsu, W. (2008). Two approaches for interaction management in timbre-aware improvisation systems. In *Proceedings of the International Computer Music Conference.* Belfast: International Computer Music Association. Retrieved from `https://quod.lib.umich.edu/cgi/p/pod/dod-idx/two-approaches-for-interaction-management-in-timbre-aware.pdf?c=icmc;idno=bbp2372.2008.062;format=pdf`

Hsu, W., & Sosnick, M. (2009). Evaluating interactive music systems: An HCI approach. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 25–28). Pittsburgh, PA.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press. doi: 10.1525/mp.2007.24.5.511

Hutchings, P., & McCormack, J. (2017). Using autonomous agents to improvise music compositions in real-time. In J. Correia, V. Ciesielski, & A. Liapis (Eds.), *Computational intelligence in music, sound, art and design* (pp. 114–127). Cham: Springer International Publishing.

Jordanous, A. (2012, 01). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation*, *4*(3), 246–279. doi: 10.1007/s12559-012-9156-1

Kalonaris, S. (2017). *Dory*: A purposively flawed and forgetful artificial musical agent. In R. Laney (Ed.), *Proceedings of the 2nd Conference on Computer Simulation of Musical Creativity.* Milton Keynes, UK: Open University.

Klein, G. A., Orasanu, J. M., & Calderwood, R. (1993). *Decision making in action: Models and methods.* Norwood, NJ: Ablex Publishing Corporation.

Lewis, G. E. (1999). Interacting with latter-day musical automata. *Contemporary Music Review*, *18*(3), 99–112. doi: 10.1080/07494469900640381

Lewis, G. E. (2000). Too many notes: Computers, complexity and culture in *Voyager*. *Leonardo Music Journal*, *10*(2000), 33–39. doi: 10.1162/096112100570585

Linson, A., Dobbyn, C., & Laney, R. (2012). Critical issues in evaluating freely improvising interactive music systems. In M. L. Maher, K. Hammond, A. Pease, R. P. Pérez, D. Ventura, & G. Wiggins (Eds.), *International Conference on Computational Creativity* (pp. 145–149). Dublin: University College Dublin.

Linson, A., Dobbyn, C., Lewis, G. E., & Laney, R. (2015). A subsumption agent for collaborative free improvisation. *Computer Music Journal*, *39*(4), 96–115. doi: 10.1162/COMJ_a_00323

Manzolli, J., & Verschure, P. F. M. J. (2005). Roboser: A real-world composition system. *Computer Music Journal*, *29*(3), 55–74. doi: 10.1162/0148926054798133

McCormack, J. (2003). *The application of l-systems and developmental models to computer art, animation and music synthesis* (Unpublished doctoral dissertation). Monash University.

Mehri, S., Kumar, K., Gulrajani, I., Kumar, R., Jain, S., Sotelo, J., ... Bengio, Y. (2016). SampleRNN: An unconditional end-to-end neural audio generation model. *Computer Research Repository*, *abs/1612.07837*. Retrieved from http://arxiv.org/abs/1612.07837

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

Moorman, C., & Miner, A. S. (1988). Organizational improvisation and organizational memory. *The Academy of Management Review*, *23*(4), 698–723.

Pachet, F. (2003). The Continuator: Musical interaction with style. *Journal of New Music Research*, *32*(3), 333–341. doi: 10.1076/jnmr.32.3.333.16861

Pachet, F., Roy, P., Moreira, J., & d'Inverno, M. (2013). Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2205–2208). New York, NY: ACM. doi: 10.1145/2470654.2481303

Peters, C., Castellano, G., & de Freitas, S. (2009). An exploration of user engagement in HCI. In G. Castellano, M. J., J. Murray, K. Karpouzis, & C. Peter (Eds.), *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots* (pp. 9:1–9:3). New York, NY: ACM. doi: 10.1145/1655260.1655269

Quigley, M., Gerkey, B. P., Conley, K., Faust, J., Foote, T., Leibs, J., ... Ng, A. Y. (2009). ROS: An open-source robot operating system. In *ICRA Workshop on Open Source Software.* Retrieved from http://www.robotics.stanford.edu/~ang/papers/icraoss09-ROS.pdf

Rowe, R. (1992a). *Interactive music systems: Machine listening and composing.* Cambridge, MA: MIT Press.

Rowe, R. (1992b). Machine listening and composing with Cypher. *Computer Music Journal*, *16*(1), 43–63.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138.

Simon, H. (1996). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Slote, M. (2004). Two views of satisficing. In M. Byron (Ed.), *Satisficing and maximizing: Moral theorists on practical reason* (pp. 14–29). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511617058.002

Snyder, B. (2000). *Music and memory: An introduction.* Cambridge, MA: MIT Press.

Solomon, L. (1986). Improvisation ii. *Perspectives of New Music*, *24*(2), 224–235.

Stark, A. M., Davies, M. E. P., & Plumbley, M. D. (2009). Real-time beat-synchronous analysis of musical audio. In U. Zölzer (Ed.), *Proceedings of the 12th International Conference on Digital Audio Effects* (pp. 299–304).

Como, Italy.

Stark, A. M., & Plumbley, M. D. (2009). Real-time chord recognition for live performance. In *Proceedings of the International Computer Music Conference* (pp. 85–88). Montreal, Canada.

Stowell, D., Robertson, A., Bryan-Kinns, N., & Plumbley, M. D. (2009). Evaluation of live human-computer music-making: Quantitative and qualitative approaches. *International Journal of Human-Computer Studies*, *67*(11), 960–975. doi: 10.1016/j.ijhcs.2009.05.007

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–402). New York, NY: Academic Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. doi: 10.1093/mind/LIX.236.433

Wallace, R. (2010). *Improvisation and the making of American literary modernism.* New York, NY: Continuum.

Wang, C., Hsu, J., & Dubnov, S. (2016). Machine improvisation with variable Markov oracle: Toward guided and structured improvisation. *Computers in Entertainment*, *14*(3), 4:1–4:18. doi: 10.1145/2905371

Weinberg, G., Godfrey, M., Rae, A., & Rhoads, J. (2008). A real-time genetic algorithm in human-robot musical improvisation. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music modeling and retrieval. Sense of sounds* (pp. 351–359). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-85035-9_24

Young, M. (2008). NN music: Improvising with a ‘living’ computer. In R. Kronland-Martinet, S. Ystad, & K. Jensen (Eds.), *Computer music modeling and retrieval. Sense of sounds* (pp. 337–350). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-85035-9_23