Fundamental considerations for empirical computational research on improvisation

Adam Linson[1]

Abstract. The field of study relevant to empirical computational research on improvisation — such as the research carried out in the Live Algorithms for Music (LAM) network — includes not only artistic and philosophical exploration of key issues related to human and machine improvisation, but also empirical research in computing, cognitive science, and psychology, which may be concerned with the make-up of (computational, cognitive, or psychological) systems involved in improvisation. This paper primarily focuses on issues relevant to the empirical study of improvisation using machines. As such, it may be of interest to researchers currently conducting empirical computational research on improvisation, but is especially aimed at those who are considering entering into this field.

1 INTRODUCTION

When considering research into an "autonomous machine improvisor", defined as the aim of the Live Algorithms for Music (LAM) network [3], there are a number of fundamental research considerations and methodological orientation points to take account of. Some work has been done on methodological issues unique to this research area, such as those pertaining to evaluation (e.g., [10] and [12]). This paper seeks to highlight further methodological issues relevant to the empirical investigation of improvisation with machines.

While we may generally pursue the study of autonomous machine improvisors, that is, artificial improvisors, we might also be concerned, more specifically, with a variety of (natural or artificial) systems involved in improvisation. In particular, since the field of research for LAM includes not only the artistic and philosophical exploration of key issues related to human and machine improvisation, but also empirical research in computing, cognitive science, and psychology, we may in these areas be concerned with the make-up of (computational, cognitive, or psychological) systems that are involved in improvisation. This paper primarily focuses on issues relevant to the empirical study of improvisation using machines. As such, it may be of interest to researchers currently currently conducting empirical computational research on improvisation, but is especially aimed at those who are considering entering into this field.

The paper proceeds as follows: the next section presents a few key examples of general approaches to computational research on improvisation. Section 3 discusses the concept of an algorithm in the present context, and considers how this concept might be relevant to methodological concerns. In Section 4, the applicability of the competence–performance distinction is considered. Section 5 addresses issues surrounding interaction and context, and also contains an extended subsection (5.1) on ELIZA, used as an illustrative example. Section 6 outlines relevant research methods, and is followed by a conclusion.

2 APPROACHES TO RESEARCH

Perhaps the most straightforwardly relevant question at the outset is that of the relation between the arts, humanities and empirical sciences in this context. Rather than a familiar "two cultures" account, it seems that one need only say that a purely artistic pursuit of an artificial improvisor must follow only an artist's prerogative — entirely independent of whether musical audiences or performers favour or disfavour the system's improvisations, or if scientists of various stripes find the system design elegant or crude. However, if a specific research question is being pursued, then depending on how it is formulated, different standards of measurement and conceptual clarifications become relevant to the study of the system. (For an in-depth treatment of important issues pertaining to interdisciplinarity, which will not be considered here as such, see [1].)

For instance, Hsu and Sosnick [10] propose that if one's goal is to increase favourable reactions to a given system (or to compare the favourability of different systems), one may seek to measure audience and/or performer reactions (for a more extensive example of such measurement, see [9]). Other work, such as research on computational creativity, may be explicitly concerned with whether specific mechanisms of a formally defined notion of creativity are being used by a system. In this setting, a case of using the correct mechanism may be judged as a successful research outcome, even if the system is viewed unfavourably by audiences and artists (see [6]). Neurocognitive research may be concerned to implement the functions of neural mechanisms with a transparent correspondence to specific biological mechanisms. Here, the empirically confirmed existence of such mechanisms in other more general research is used to determine the success of their models and architectures (for an example, see [2]).

3 WHY ALGORITHMS?

One way we can begin to identify fundamental considerations of this research is by looking at a correlate to the use of the term "algorithm" in the concept of "live algorithms for music". Although the term is used in LAM to describe the high-level behaviour of improvisation, there is nevertheless a resonance with Marr's [15] "three levels" of information processing, which he formulates in relation to research on vision (although they enjoy a wider application in contemporary science). The three levels described by Marr are "computational theory", "representation and algorithm", and "hardware implementation". The "hardware implementation" for Marr corresponds to physiology (with respect to vision, in the context of his

[1] Open University, UK, email: adam.linson@open.ac.uk

work). In the case of an artificial improvisor, not only the machine hardware, but also the software environments would be represented at this level. For example, for a given system, was a laptop used or an Arduino? What RAM, CPU, chip architecture, etc., was used? When capturing input from a human improvisor, were microphones or pick-ups used, and how were they set up? And so on. Questions posed at this level might be of some interest to the LAM community. However, following Marr's conceptualisation, one could say that a given artificial improvisor could be implemented, for example, in different software languages or with different hardware, while retaining its identity as a specific system design.

In research on machine improvisation at the opposite end of Marr's spectrum, "computational theory", we would likely expect high-level descriptions of improvising, for example, those found in the LAM description, such as the specification that the system "interacts with human performers by listening to contributions in its musical environment, preparing musical material that has at least an element of originality and appropriateness, and delivering this material back into the environment" [3]. There are also more specific concepts that are relevant at this level of abstraction, for example, that the system should "know" how to make certain decisions: "when to play or not, when to modify activity in any number of parameters (loudness, pitch, tone quality), when to imitate or ignore another participant, when to 'agree' the performance is concluding", and beyond that, "when to make a decision. And why" [4].

The question that seems most central to this research is: how does a system "know" (or how does one get it to know) what is necessary for it to participate in human–computer improvisation? What *algorithms* constitute the system's "decision" to play or not? This characterisation seems to fit with Marr's second level, that of "representation and algorithm", (here) the question of how the high-level theory of improvisational decision-making can be implemented, using what input and output, and what transformations (see [15, p. 24ff.]). While this level of understanding — the algorithmic level — is suggestive of low-level mechanisms, certainly in Marr's original usage, it can be extended to the consideration of high-level general behaviours such as musical improvisation. This conception will help to delimit the scope of how to approach empirical research on live algorithms for music.

4 COMPETENCE VERSUS PERFORMANCE

Following the general conceptualisation borrowed from Marr, we might move on to consider the idea of an improvisational competence. The distinction raised above between an improvisor and an improvisation system becomes acutely relevant here. If the question is, What constitutes the competence of a musical improvisor?, we might say that this question is formulated at the computational level, and that it can be explored using a system that implements the computation using any variety of algorithms. On the other hand, we might ask, What constitutes the competence of a musical improvisation *system*?, which we might describe at either the computational or algorithmic level. If we describe competence at the algorithmic level, we may seek to determine if a given hardware (and software) implementation is the best way to realise a concrete instance of a functional system. The latter question may seem more trivial, but it is always lurking in the background of the former question: in the sometimes messy practicalities of research, one could say that a competence theory (at the algorithmic level) has not been invalidated just because some specific implementation (at the hardware level) did not perform to the expected standard. Consider that one may have a good theory of what constitutes improvisational competence at the algorithmic level, but that the specific hardware or software used in the implementation may not be sufficient to perform as expected in real time.

Moreover, some researchers may seek to uncover inherent advantages or limitations in choosing among Max/MSP, Pd, Csound, ChucK, etc., and choosing exactly how specific algorithms are implemented in a given environment, using this or that native function, integers or floating points, etc. One may also explore differences between using a 64-bit multicore processor or an Arduino, or multiple Arduinos, or a portable electronic device, or a remote server (and so forth), guided by ideas about why one might favour one realisation over another. Here, empirical research could be used to determine that the performance with a specific implementation best realises a given competence.

A further issue concerning the competence–performance distinction must be raised. If one defines improvisational competence at the computational level, does one hold that an isomorphic mechanism (or architecture) is present within the algorithm, or simply that an assessment of the algorithm's performance is expected to yield an approximation of the competence, for example, as an emergent property? To clarify this distinction, consider Dreyfus' [8] example of the competence of balancing while riding a bicycle:

> A man riding a bicycle may be keeping his balance just by shifting his weight to compensate for his tendency to fall. The intelligible content of what he is doing, however, might be expressed according to the rule: wind along a series of curves, the curvature of which is inversely proportional to the square of the velocity. The bicycle rider is certainly not following this rule consciously, and there is no reason to suppose he is following it unconsciously. Yet this formalisation enables us to express or understand his *competence*, that is, what he can accomplish. It is, however, in no way an *explanation* of his *performance*. It tells us what it *is* to ride a bicycle successfully, but nothing of what is going on in his brain or in his mind when he performs the task. [p. 190, original emphasis]

Following this example, one could imagine two scenarios: one with a robot that in fact uses the abstract formalised competence directly, either explicitly or implicitly, such that one could in some way "point" to where in the machine this competence is accounted for (even if it is distributed among components). But in another scenario, a differently constructed robot may achieve this balancing act using an assemblage of components that in no way contain an (explicit or implicit) mechanism that stands in an isomorphic relation to the abstract formalisation. Instead, when the system balances successfully, we say that its performance (like that of the human in Dreyfus' example) can be ideally represented according to the formalisation, but that the formalisation in fact describes an emergent property of the system — there is nowhere to point to inside the system where the formalisation is isomorphically represented.

These senses of competence (or at least roughly similar ones) are discussed by Clark [5], though he is concerned with the explanatory role of competence in the context of different theories of cognition. Arguably, his description of the relation of a competence to a connectionist model is similar to the conception of competence described by Dreyfus above, where the competence description need not be represented explicitly or implicitly in the system. When formulating a research question for the study of artificial improvisation systems, one must be clear about exactly what type of competence is being considered.

5 INTERACTION AND CONTEXT

Another point to consider is that, in the human–computer interaction, one can abstractly treat an artificial improvisor as a whole, and look at a single point of interaction between human and system. One can also consider the relation of different subsystems to each other and to different aspects of the interaction. This consideration is especially pertinent to the topic of emergence. Specifically, there may be an emergent 'local' behaviour that arises from the interaction between two mechanisms in an improvisation system (for example, a system's apparent ability to harmonise may result from several subsystems, as with Rowe's *Cypher* [17]). One can also perceive an emergent 'global' behaviour of such systems, which, similarly to how the bicyclist is described, can be addressed according to a general model of (in this case) improvisational competence. One can also expect to encounter emergent dynamics in the human–computer collaboration, such that certain properties of the performance emerge from the interaction between human and computer. Emergent properties are relative to an interaction context, and thus the notion of emergence must be used with care for it to be helpful in descriptive or explanatory accounts.

Here, we should also bear in mind a distinction between the smaller and larger systems involved [11]. This is an especially important point for interactive systems, because it needs to be clear exactly what constitutes the system being investigated. Are we looking inside a narrowly construed computer system (i.e., software and hardware) and describing how, for example, a "decision to play" is set up? Or, are we looking at an entire system of elements that includes the software, the computer running the software, the relation between elements external to the computer that may include humans, computers, other devices, spaces (e.g., a stage, a studio), and so on? To take but one significant example, for an artificial improvisor that interacts with multiple musicians simultaneously, the question of how it responds is importantly altered depending on whether multiple sound sources are combined (mixed) prior to a single computer input, or if they are received by the computer as independent inputs — and in the latter case, if and how the signals are combined internal to the system.

Similarly, there is the question of how the semantic context surrounding the interaction is set up, and how this may bear on the situation. At the broadest level, there are cultural norms in both music and science that affect how a given human–computer musical interaction is received and legitimated, how the resulting music itself is received aesthetically, and how the research setting accords a standard of rigour. When these semantic contexts overlap, as they typically do, it is important to disentangle them, reach a critical understanding of the conflicting positions involved, and focus on what is relevant to a given research question, which must also be justified by an argument.

5.1 ELIZA

An important case related to the issue of semantic context is Joseph Weizenbaum's ELIZA program [19, 20], designed to verbally interact with a human participant. ELIZA is based on pattern matching and rule-based analysis and generation, an approach that can also be found in some musical improvisation systems. But rather than comparing design details, I wish to compare here the interactive contexts of research on ELIZA and research on artificial improvisors. While this example is nearly fifty years old, it provides a striking analogy to the present context, which I will illustrate here in depth.

For ELIZA, there are several factors to consider that pertain to the interaction context. For one, the system was not claimed to be doing anything like what humans do, so we are not confronted with a case of a neurocognitive model being validated by empirically identified mechanisms, nor even of a formalised model of grammar such as those typical in contemporary linguistics (another area where the competence–performance distinction arises). Nevertheless, ELIZA is a system that apparently performs interactively with humans reasonably well. Or does it? The claim must be further delimited to be understood correctly. If the semantic context for a human–computer interaction with ELIZA is the general task of having a verbal conversation, then in fact it does not perform well at all. Yet, when the interactive context is further narrowed to a specific kind of psychotherapy, the system does appear to perform up to expected standards. Thus, similarly, for a system that performs interactive music, one must consider how specifying (for instance) "free improvisation" as the mode of musical engagement structures the interpretation and evaluation of the system performance.

Weizenbaum points out that framing the system performance for participants, by influencing their expectations, significantly affects their interpretation and evaluation of the system behaviour. In the case of ELIZA, the human participant is led to expect a specific type of conversation; in the case of an interactive free improvisation system, for example, the participant is led to expect a free improvisation. In both cases, participants interpret the system's behaviour within a specified, delimited context.

In the original descriptions of ELIZA [19, 20], a conversation is presented as something that can be intuitively recognised, without the requirement of a specific experiment or method of evaluation. Example conversations with ELIZA (actually, with its DOCTOR variant) are given as evidence that the system is fit to hold a conversation. However, rather than any variety of conversation, the verbal interaction is framed in a certain way for the human interlocutor, namely, as one would talk to a Rogerian psychiatrist, meaning that a speaker is encouraged to continue speaking in response to the psychiatrist's prompts. Weizenbaum gives an example of an exchange with ELIZA in which a human participant mentions a boat, eliciting the system response, "tell me about boats"; such responses, although simplistic, serve to direct the conversation and keep it going. According to Weizenbaum, this allows for the system to converse without recourse to any knowledge of the world.

Weizenbaum uses the example of ELIZA to raise questions about the notion of machine understanding, just as we might use research with machines to explore questions about improvisation. With respect to ELIZA's internal mechanisms, it clearly does not use language in the same way that humans do, and generally lacks a technical apparatus that would allow it to have even the possibility of understanding. But it is important to emphasise that, from the perspective of the human interacting with the system, the performative act of conversing can take place, analogously to how humans might converse. In other words, the system provides the human participant with a certain kind of interaction, and this interaction can potentially become, for the participant at least, what is termed a "plausible" conversation (presumably meaning plausibly realistic or human-like). For a human–machine improvisation, we might instead consider what makes for "a meaningful experience of interacting with a machine intelligence" and what allows us to be able to "explore novel musical spaces" with such a machine [3].

Regarding conversational context and interpretation, Weizenbaum [20] states that:

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

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

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

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

Building upon this approach to theory, system design, and experimental evaluation, rather than investigate the more general problem of machine improvisation, which does not suggest a clear standard of measurement, we may instead seek to reveal insights about improvisation itself, or perhaps about improvisational interaction between humans (without machines). As with ELIZA, one approach to such research is to take a typically human–human interaction, and substitute one human with a computer that has the desired competence for the interaction.

6 RESEARCH METHODS

Once a research question has been determined, there are familiar methodological issues to respond to concerning the notorious quantitative–qualitative divide. Davidson [7] offers a well balanced account of the variety of quantitative and qualitative empirical research methods that address music as social behaviour, which certainly may be one of the aims of research on improvisation with interactive systems. One may also or instead be concerned with other aspects of improvisation that might be independent of social behaviour.

One must consider the relevance of different types of data to specific research questions, and one must also consider the research value of the data. For example, a questionnaire or structured interview might be useful for getting large sample sizes of participant feedback, but then one must be careful with respect to determining statistical significance. With a small sample size, it might be more valuable to get the type of rich data that one can obtain with a semi-structured interview, but then one must face the trade-off of increased time and effort when it comes to analysing the data.

Wengraf [21] writes of how semi-structured interviews typically require more preparation and more analysis than fully structured interviews. This is because in a semi-structured interview, one must closely pay attention to the answers, understand them in the context of the interview and the study as a whole, and engage in a more conversational role with the participant. In contrast, in a fully structured interview, one proceeds through a series of questions without any regard for the content of the answers during the interview. Wengraf [21, p. 25] states that, as an interviewer in the semi-structured mode, "I both respond to what has been said (and not said) so far in the conversation, but also act in the present in anticipation of possible futures of the conversation, which I wish to move towards (or to avoid)". This type of conversation is "much more artful" than is suggested by turn-by-turn or question-and-answer models; sometimes, for example, an "interviewer goes back to [a] 'dropped hint' at a much later point and gets a different dimension [of the topic] that was present but not observed at the time" [p. 41].

Assuming one gets the appropriate data for one's research, one must also best determine how to appropriately analyse the data. For quantitative research, statistical methodologies are of paramount importance, and one must, for example, determine if Bayesian methods will be used, how outliers will be identified, etc. There is a vast number of qualitative data analysis methods, such as interpretive phenomenological analysis (IPA), Framework, ethnography, grounded theory, etc. But as Ritchie and Lewis [16, p. 201] state of qualitative analysis, while there are many different traditions and methods, with different notions of what should comprise the main assumptions and focus of analysis, "distinctions are not always clear cut", and boundaries are often crossed between traditions, or within individual studies. One must not only demonstrate *that* one is using a systematic methodology for one's research, but *why* a particular method (or hybrid of methods) is appropriate.

It should also be noted that when the object of study is participant experience, the everyday broad construal of experience must be narrowed in order to be rendered manageable. For example, experiential studies on improvisation (without computers) have been conducted without a connection to a specific performance (e.g., [14]), or have used listening to recordings as a means for improvisors to reconstruct internal mental narratives of their performance (e.g., [18]). The author's research [13] has used performers' verbal feedback about their experience performing with an artificial improvisor directly following the performance, to elicit a more immediate recollection. Here, approaches to investigating experience from the humanities and social sciences, for example, ethnography, might be well suited to particular research questions, whereas for other research, it might be decided to examine physiological properties of experience that may not be accessible to an 'in the wild' observer, or may not be revealed by self-reporting (e.g., excitement measured by a rise in heart rate).

7 CONCLUSION

This paper has outlined some fundamental considerations for empirical computational research on improvisation; there are undoubtedly more that have not been covered here. Generally speaking, it is clear that, on one hand, this field provides an interesting platform for both artistic experimentation and philosophical research. In these contexts, a familiarity with improvisation (as a listener or performer) might serve as a point of entry that can ultimately lead to valuable insights in terms of aesthetics, ethics, social and political theory, and so forth, which should continue, and should be further encouraged. On the other hand, this intrinsically multi- or interdisciplinary research is clearly not restricted to the arts and humanities. Empirical researchers might also make a valuable contribution to the study of improvisation using machines, and their efforts may extend beyond the confines of traditional computer science. For those considering embarking on empirical research in this field, it is hoped that the above considerations will serve as a useful basis for orientation.

REFERENCES

[1] Andrew Barry, Georgina Born, and Gisa Weszkalnys, 'Logics of interdisciplinarity', *Economy and Society*, **37**(1), 20–49, (2008).

[2] Harold Bekkering, Ellen RA De Bruijn, Raymond H Cuijpers, Roger Newman-Norlund, Hein T Van Schie, and Ruud Meulenbroek, 'Joint action: Neurocognitive mechanisms supporting human interaction', *Topics in Cognitive Science*, **1**(2), 340–352, (2009).

[3] Tim Blackwell and Michael Young. Live algorithms: Tenth anniversary meeting. `https://sites.google.com/site/livealgorithms/`.

[4] Tim Blackwell and Michael Young, 'Live algorithms', *AISB Quarterly, Issue 122: Autumn 2005*, 7–9, (2005).

[5] Andy Clark, 'Connectionism, competence, and explanation', *The British Journal for the Philosophy of Science*, **41**(2), 195–222, (1990).

[6] Simon Colton and Geraint A Wiggins, 'Computational creativity: The final frontier?', in *Proceedings of ECAI 2012*, eds., Luc De Raedt et al., pp. 21–26, (2012).

[7] Jane W. Davidson, 'Music as social behavior', in *Empirical musicology: Aims, methods, prospects*, eds., E. Clarke and N. Cook, 57–75, Oxford University Press, Oxford, (2004).

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

[9] Arne Eigenfeldt, Adam Burnett, and Philippe Pasquier, 'Evaluating musical metacreation in a live performance context', in *Proceedings of the Third International Conference on Computational Creativity*, pp. 140–144, (2012).

[10] William Hsu and Marc Sosnick, 'Evaluating interactive music systems: An HCI approach', *Proceedings of New Interfaces for Musical Expression (NIME)*, 25–28, (2009).

[11] M.A. Jackson, *Problem frames: analysing and structuring software development problems*, Addison-Wesley, Boston, 2001.

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

[13] Adam Linson, *Investigating the cognitive foundations of collaborative musical free improvisation: Experimental case studies using a novel application of the subsumption architecture.*, Ph.D. dissertation, Open University, 2014.

[14] R.A.R. MacDonald and G.B. Wilson, 'Constructions of jazz: How jazz musicians present their collaborative musical practice', *Musicae Scientiae*, **10**(1), 59–83, (2006).

[15] David Marr, *Vision*, W.H. Freeman and Company, San Francisco, 1982.

[16] J. Ritchie and J. Lewis, *Qualitative Research Practice: A Guide for Social Science Students and Researchers*, SAGE Publications, London, 2003.

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

[18] M. J. Sansom, *Musical meaning: A qualitative investigation of free improvisation*, Ph.D. dissertation, University of Sheffield, 1997.

[19] Joseph Weizenbaum, 'Eliza: a computer program for the study of natural language communication between man and machine', *Communications of the ACM*, **9**(1), 36–45, (1966).

[20] Joseph Weizenbaum, 'Contextual understanding by computers', *Communications of the ACM*, **10**(8), 474–480, (1967).

[21] T. Wengraf, *Qualitative Research Interviewing: Biographic Narrative and Semi-Structured Methods*, SAGE Publications, London, 2001.