

Musical and Meta-Musical Conversations

Ben-Tal (Kingston University) and David Dolan (Guildhall School of Music and Drama)

Abstract

Improvising music is a particularly challenging task for a human musician, not to mention AI systems. It requires proficiency in many facets of musical behaviour simultaneously: listening, imagining, enacting. Successfully improvising with other musicians requires fusing several layers of musical creation: shaping sounds and generating surface level details, linking those to larger structural units such as phrases and sections, and creating convincing narratives or trajectories. And doing so in real-time while maintaining connection with partners. In this paper, we describe an AI system that improvises music with a human performer in tonal context though not restricted to specific styles or genres. We discuss our methodology in developing this system, and the importance of slow and deliberate refinement for a successful application of novel technology for meaningful music making. We structure this paper as a conversation between the authors -- showcasing the different perspectives on joint music making, and the ways these perspectives shape the developing system.

Prelude

This collaboration emerged out of informal conversation between the authors about improvisation. Ben-Tal is a composer/researcher who has been using Music Information Retrieval (MIR) techniques and AI as tools for composition. Author 2 is a performer/improviser and researcher on improvisation, creativity and expressive performance with little knowledge of music technology. Dolan became intrigued but also highly sceptical about Ben-Tal's ideas of musical dialogues between human and computer as a basis for co-creation. They agreed to meet and trial the possibility of real-time improvisation between piano and computer. By his own admission, Dolan came to this first session assuming he will prove the inadequacy of such a set-up for joint improvisation based on extended tonal music idiom. He found himself equally surprised and alarmed when he experienced moments that felt, to himself, as real dialogue with the machine. This proof-of-concept session provided the starting point for an ongoing collaboration: developing a unique duo-improvisation within the context of computationally creative tools, real-time interaction, tonal-music and human-computer interaction. Central to this work are musical dialogues between Dolan on the piano and Ben-Tal's computing system as they improvise together. These are surrounded and complemented by conversations between the authors about the system, about improvisation, composition, performance, music and AI.

This presentation starts from a description of the current improvisation set-up and the development that allowed us to arrive at this stage. The following section re-enacts some of the conversations that the Authors engaged in, in order to illuminate the learning and discovery process they underwent together. We will end by drawing out important themes emerging from the musical and meta-musical conversations in relation to current debates around music and AI.

Exposition

First Theme – Artificial Improviser

We describe the role of the computer in these performances as an artificial improviser even though the term has no clear definition and the range of approaches used over the years is wide ranging¹. Some systems aim to work within a particular musical style². Others operate as an extension of an improviser's own practice³. This artificial improviser was developed by Ben-Tal with the explicit intention of creating a musical dialogue between an AI system and Dolan. Two fundamental design decisions informed the development of this artificial improviser. First, Ben-Tal is a composer: he did not want to develop a musical instrument. His preference was for a system that handles the moment-to-moment sound production automatically. His role – on stage – would be as a supervisor or executive: monitoring the system's behaviour, influencing its responses but without the need for constant action. The second important design decision was to try and minimise hard-coded assumptions on the input. Most improvisation systems include such assumptions which impose restrictions on the human performer. Somax⁴, for example, uses tempo when analysing the corpus and the audio during a performance. The performer can set the tempo but is still restricted at any given moment. The Reflexiv Looper⁵ and Shimon The Robot⁶ both need a lead sheet as a frame when accompanying a human performer. While Dolan improvises within the realm of expanded tonality, the basic conventions of tonality are not coded. The computer does not look for a regular pulse, major-minor chords, cadences, or other stylistic conventions.

The initial development phase of the system used recordings made of Dolan improvising solo on the piano. Using these as simulations, Ben-Tal developed a modular system that

¹ Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., & d'Inverno, M. (2018). Computational systems for music improvisation. *Digital Creativity*, 29(1), 19-36.

² For example: Ramalho, G. L., Rolland, P. Y., & Ganascia, J. G. (1999). An artificially intelligent jazz performer. *Journal of New Music Research*, 28(2), 105-129; Thom, B. (2000, August). Artificial intelligence and real-time interactive improvisation. *Proceedings of the AAAI-2000 Music and AI Workshop* (pp. 35-39); Dannenberg, R. B., Gold, N. E., Liang, D., & Xia, G. (2014). Methods and prospects for human-computer performance of popular music. *Computer Music Journal*, 38(2), 36-50.

³ For example Lewis, G. E. (2000). Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal*, 10, 33-39 or Rowe, R. (1992). Machine listening and composing with cypher. *Computer Music Journal*, 16(1), 43-63.

⁴ Borg, J. (2019) Somax 2: A real-time framework for human machine improvisation. *Ircam, Tech. Rep.*

⁵ Pachet, F., Roy, P., Moreira, J., & d'Inverno, M. (2013, April). Reflexive loopers for solo musical improvisation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2205-2208).

⁶ Hoffman, G., & Weinberg, G. (2010). Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3097-3102).

uses machine listening techniques and offers a range of responses⁷. The computer uses MIR tools to extract pitch information (chromagrams), rhythmic information (inter onset intervals; IOI), and some timbral features (spectral centroid, perceptual loudness) from the piano's acoustic signal. The computer isn't trying to fit rhythmic information into predefined categories (pulse; duple/triple meter; bars; phrases) which would require hard-coding assumptions. Instead, the Kernel Density Estimation method (KDE)⁸ is used to construct a statistical distribution of the IOI values. The peaks of this distribution are used to estimate the prevalent rhythmic categories currently used by the pianist. Each new IOI value is compared to the current categories derived from past values. If too many incoming values no longer fit close enough to the current categories the computer runs KDE on the list of IOIs to update the list of time intervals. Notably, this approach does not provide metric information such as bars, weak and strong beats, pulse and subdivisions. Which means the system does not have access to such factors even though they are important in tonal music.

Chromagrams are used as a proxy for harmony – again an oversimplification compared to the deep knowledge of the language of tonal music that human musicians possess, but a simplification that avoids the need for complex analysis (difficult in real-time) and for hard-coded assumptions (e.g. look for triads; relate harmony to some putative tonic). The strongest *n* notes of the Chromagram are collected every second (Author1 sets the value of *n* dynamically in the performance). This collection is the basis for generating much of the pitch material in the sounds the computer contributes to the dialogue. The system also uses the Prediction by Partial Match⁹ method to estimate predictability/surprise of the current event.

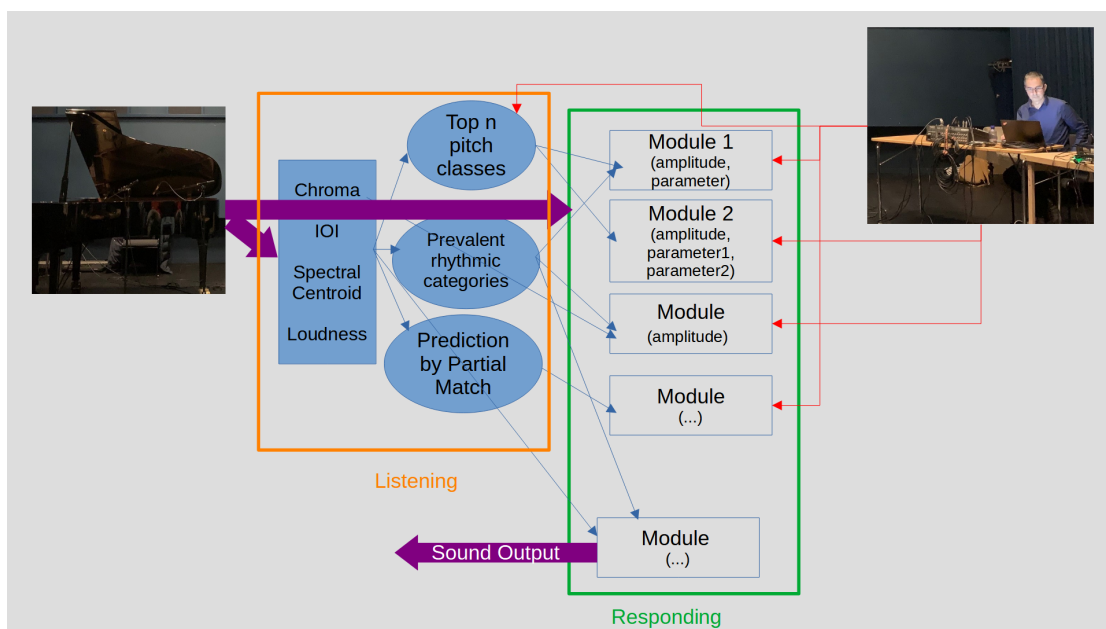


Illustration 1: Schematic outline of the system. Listening component extract features and

⁷ Building on previous work by Ben-Tal. See Ben-Tal, O. (2023). Weak interactions, strong bonds: Live electronics as a complex system. *Organised Sound*, 28(1), 122-132.

⁸ Węglarczyk, S. (2018). Kernel density estimation and its application. In *ITM Web of Conferences* (Vol. 23, p. 00037). EDP Sciences.

⁹ Whorley, R. P., Wiggins, G., Rhodes, C., & Pearce, M. T. (2010). Development of techniques for the computational modelling of harmony.

analyses them. independent modules use this data to produce responses. Ben-Tal adjusts some parameters during the performance.

The sound production is handled by a growing collection of independent modules. In performance Ben-Tal activates and deactivates these modules. At any given moment any number - usually 3-5 - will be active. In some moments only one module is operating, but as many as 8 have been used simultaneously in performance. Most modules are defined by (1) their use of the data extracted from the performer (2) an internal musical-logic that defines their dynamic responses. Below is the outline of one such module:

1. take the list of rhythmic categories using the current estimate
2. append a rhythmic value 1.5x time the longest value
3. append a rhythmic value of 0.1 seconds ('grace note')
4. Randomly select a pitch from the n-strongest bins in the Chromagram. (according to the current value of n set by Ben-Tal)
5. Use a plucked string physical model to synthesise a note with the selected pitch.
6. Use weighted random from the rhythmic value list created in steps 1-3 for the IOI to the next note.
7. repeat

In this case both the rhythm and the pitch parameters are derived from the current piano context. The algorithm was shaped through an iterative process of listening to the output, in conjunction with the recorded improvisations, until Ben-Tal was satisfied with the result. For instance, initially the generated patterns lacked shape and sounded like idle noodling. The addition of the longer and shorter IOI categories created local gestures which added focus and shape. The timbral qualities of the synthesis, the weighting of the rhythmic value choice and the octave transposition from the Chromagram to the synthesis are similarly important factors that shape the result and were derived gradually through the process. In other words, the musical-logic of the algorithm encodes Ben-Tal's compositional aesthetics in dialogue with Dolan's musical idiom.

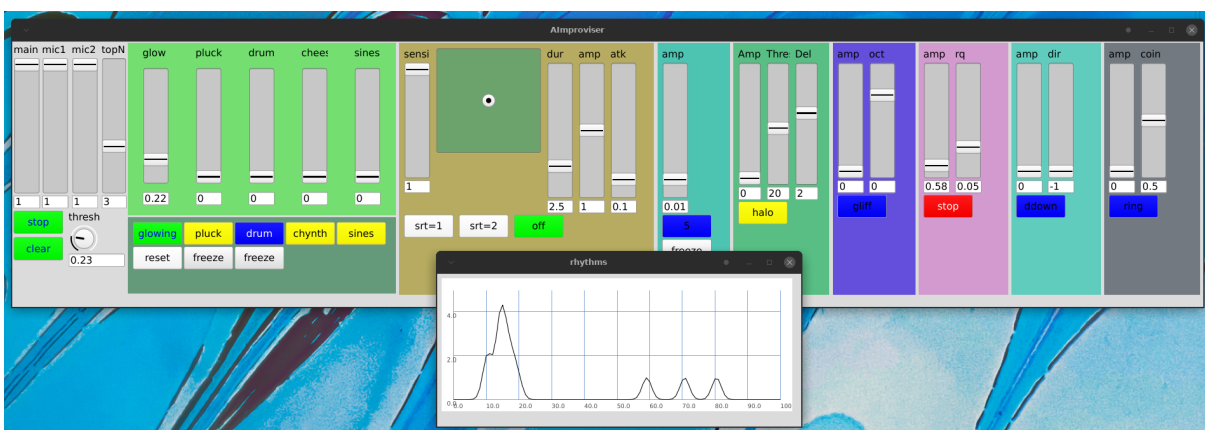


Illustration 2: interface windows used by Ben-Tal: controls at the top, histogram of rhythm below

Second Theme – Human Improviser

If we consider music as a combination of the 'What' dimension - compositional content - with the 'How' dimension - shaping and interpreting this content - improvisation is the fusion of both dimensions. In real-time. The 'What' dimension includes the creation of melody, harmony, gestures, phrases, texture, etc. The 'How' dimension encompasses the real-time choices regarding performance-related parameters (e.g. dynamics/intensities, durations/timing/rhythms [rubati], changes of timbre and elaborations of thematic elements). Fusing both dimensions at the same time requires drawing simultaneously from different pools of know-how, and dealing with long-term structures (phrases, sections, movements' overall form) as well as short-term ones (motives, gestures). While performing improvisations, Dolan functions at several structural levels/layers simultaneously, including awareness of directionality (from departure points to goal-points) within larger time units and dealing with shorter-term expressive gestures. In his structural thinking he relates to deep structural pulses of longer durations (whole-bar or double-bars, for example) underlying shorter time units and values above these inner pulses¹⁰. For example, deeper harmonic movements frame motivic gestures on the foreground level. These inner pulses function as tactus, but instead of being fixed, they can grow longer or shorter according to the improvising performer's needs in relation to what happens on the surface (the actual fully elaborated content performed).

The art of improvisation has been an integral part of European music-making culture up until the end of the 19th century. This, both while performing repertoire (elaborated/embellished repeats of themes, eingang & fermata points, cadenzas and interludes) and improvising fantasies, preludes and other forms independently of specific repertoire works¹¹. While this phenomenon was mostly applied in the context of solo performance, Dolan's method and practice place significant emphasis on ensemble improvisation in various styles and aesthetic contexts, including Baroque, Classical, Romantic, post-Romantic, extended tonality, and tonally-free stylistic languages.

Improvising within such a tonal, modal, or extended tonal idiom requires the performer to adhere to stylistic/aesthetic constraints. This has important implications to the recently established link between improvisation, characterised by spontaneous responses to the unexpected¹², and states of flow¹³. The spontaneous aspect of behaviour most of us know from daily life is one of the elements in the teaching and practice method developed by Dolan, which focuses on fusing real-time spontaneous creative decision-making with a deep assimilation of knowledge that allows for its application in real-time, while in the flow of performing the improvisation. This could be described as an Improvisational state of mind. When improvising with others the twofold enhanced challenge is of keeping flow, expression

¹⁰ Pressing, J. (1988). *Improvisation: methods and models. Generative processes in music*, 129-178.

¹¹ Gooley, D. (2018). *Fantasies of improvisation: Free playing in nineteenth-century music*. Oxford University Press.

¹² Dolan, D., Jensen, H. J., Mediano, P. A., Rajpal, H., Rosas, F., & Sloboda, J. A. (2018). The Improvisational State of Mind: A Multidisciplinary Study of an Improvisatory Approach to Classical Music Repertoire Performance. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01341>

¹³ Csikszentmihalyi, M., Csikszentmihalyi, I. S, Ed; (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge University Press.

and coherence of oneself within relevant aesthetic/stylistic references while integrating partners' contributions. How could a mindless machine operate under these conditions?

Fugue¹⁴

Ben-Tal (A1): How was the process of improvising with the computer different to developing improvisation with a new human performer?

Dolan (A2): Improvising, or indeed performing, with human musicians, benefits from various extra-musical or quasi-musical cues: the way they breath and move, body language and facial expressions, etc. These are missing from the exchange with the computer. Which makes the fact that there is a meaningful musical exchange between me and that 'thing' (the computer) feel somewhat awkward or strange. Another major difference is the sonic landscape when the sounds I hear are all coming from the speakers not from an acoustic source.

A1: This relates to one of the differences between your piano and my computer. The piano has an identity as a mechanical source for sound making. It offers a limited range of sounds. In contrast the computer has few limitations. The speakers, as a sound source, is one limitation and maybe processing power to a degree. But anything else is down to my choices and abilities to imagine and implement sound generation processes.

A2: Indeed, but there is another dimension to this identity of the instrument which is that when I am improvising as part of an ensemble. With a string quartet, for example, I become a part of a piano quintet. I am aware how the piano is part of this ensemble and it's a slightly different instrument then when I am improvising solo or with another instrument like the trombone or the oud.

A1 It is interesting that you mention both a conventional ensemble like the piano quintet as well as improvising with an oud player. The long history of the former informs the improvisation and provides context for it – for the players as well as the audience. But with an oud player you don't have this context to draw on.

A2 that is true, and the main difference between these was in terms of musical language. With the oud player, we wanted to explore the possibility of our individual musical idioms together: a kind of intercultural exchange. So the challenge, for me as a pianist, was to dialogue with music grounded in classical Arabic art-music language. While with this 'thing' – your AI programme – the musical dialogue is flowing in one combined language. Which is what I found so surprising and somewhat uncanny.

A1 can you explain what you mean?

A2 I hear and feel a dialogue but I don't know who exactly am I in dialogue with. And I don't understand how this sense of dialogue is achieved. When I perform with others the shared experience - musical, expressive, emotional - is a large part of it. But here there is no shared

¹⁴ Two short performance extracts are available online: [Example1](#), [Example2](#)

experience. Yet there is, somehow, shared music making. Which is uncanny almost in the literal sense of the word: unknown and possibly unknowable.

A1 The system is designed for musical dialogue. It tries to listen to you and produce musical responses that are appropriate to the context. Appropriate is, of course, based on my aesthetic choice as a composer. So the system encodes an approach to musical listening as well as aspects of my compositional practice. At the same time it was designed to work with *you* – I developed it with your improvisations in my mind and my ears. It is tailor made. Which leads to another question: do you find that you need to listen to the system differently compared to human performers?

A2 It is difficult to pinpoint exactly how it is different, but it is. I noticed that there are less gradations or continuous shaping of tone/sound, timbre and dynamics generated by the computer compared to human musicians playing their instruments or singing.

A1 perhaps this is related to the modularity in my system with individual components which I activate and deactivate. So each activation brings in new sounds or new musical ideas. Is that what you are referring to? Or is it even with individual processes – which mostly have their distinct timbres – where there is less shaping of ‘notes’ and gestures?

A2 The “otherness” I am talking about relates also to individual processes, although the “entry” and “departures” of processes are obviously a significant part of the phenomenon. I think that when it comes to individual processes, the evolution of sound and the relations between how curves of intensity, timbre, pitch and durations unfold are different compared with acoustical instruments. At the same time your system is also not responding well to dramatic changes on my part. Unlike a human partner which would recognise a drastic shift into a new section.

A1 That makes sense. My system ‘listens’ to the local context: the pitch content, the time intervals between onsets, some timbral aspects. What it still doesn’t do is listen to longer time-spans and try to identify these contours you describe. This problem of structure in generative music (more accurately, generating music with coherent or plausible structure over longer durations) is not unique to my system¹⁵. And automatic segmentation of music - including identifying these section boundaries - is another perennial challenge even when processing off-line¹⁶. The processes I designed do have local structure - I think you refer to it as directedness - a shaping of the sounds and the events generated into gestures or patterns with focus. But the larger scale shaping is done by me and I only have crude controls at the moment, not nuanced continual shaping.

A2 Perhaps related to that, compared with performing with human musicians, I find myself more surprised by what I hear coming out of the speakers and I am not sure it is just because of the more varied timbres. I worked with extraordinary and very versatile musicians but this AI system “manages” to surprise me more often.

¹⁵ e.g. Dai, S., Jin, Z., Gomes, C., & Dannenberg, R. B. (2021). Controllable deep melody generation via hierarchical music structure representation. Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR, Online, November 7-12

¹⁶ Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668-696.

A1 Is this part of what you find rewarding about this collaboration?

A2 In part. Initially it was curiosity – I generally can't resist new openings. And I find some of the sounds you create inspiring. After a few sessions together, an additional rewarding aspect is that the good moments become better each time. You remember that at one point out of nowhere I started a fugato texture. This is not something we tried before and yet your system followed me and contributed an appropriate fourth part to the texture, which felt like magic.

A1 Yes, that was a wonderful and very unexpected moment. My impression was that even you didn't know you would embark on such a Baroque-style prelude with a complex three-part counterpoint. And yet the system followed you. Part of it was because some of the processes I implemented transform, in non-trivial ways, the sounds of the piano. And as your texture became louder and denser the transformed version also increased in similar ways thus matching you. But the synthesised responses - where new sounds are generated by the computer based on pitch-time data extracted from your performance - also worked. There is enough connection in the way I designed the machine listening part and the way it recombines this information to produce new, but related, material.

A2 At the same time I also find it difficult to understand why this dialogue sometimes works but at other times it doesn't. As if I and that 'thing' are at cross purpose. And it goes back to this uncanny feeling. With other improvisers I can feel, in the moment, when one of us is not sufficiently attuned and I know what I can do to regain this tuning-in. But not with the AI.

A1 This does sound like the famous uncanny valley - we find it difficult to deal with an AI that appears too human in some respects while failing in others. Perhaps the relative success of the system to respond to you raises your expectations from it - expectations it sometimes fails to meet. In some ways working with this AI is neither like working with another highly experienced and proficient human improviser, nor is it like the relationship you have with your students. It has elements of both and elements that are unique to the system.

A2 In some ways it is like a student in that I notice improvement from session to session and I wonder how much is down to you fine-tuning the code between sessions, how much is down to what you do live in the performance, and to what extent I am learning - not necessarily in a conscious way - to adapt to the patterns that the computer creates.

A1 I do make small changes in the code between our sessions based on what happens in them. It could be about the sound qualities and it could be about the way the local processes work - what I described as the musical-logic programmed algorithmically. And our discussion about how you construct your improvisation or how you conceptualise the music informs those changes. Not that I am trying to mimic your approach but I try to find ways of integrating the electronics even better. But another element of this is the way we listen to each other. You learn to listen and respond to what you hear and I, in turn, learn how the computer responds to your input and how these sound together. I will give you an example. One of the modules I designed grabs a snippet of your sound, and then starts a short process of repetition and transformation of it. And it has been waiting there in the code and I rarely used it because I didn't actually like the way it sounded. And then you were playing a motif of staccato notes and I realised this should work together. What I heard coming out

when I turned this process on at that moment was nothing like what I heard when I developed this code but it worked. Sonically and musically. So it surprised me too though, when I turned this process on at that particular moment, I was expecting a good surprise.

A2 am I right in saying that you are learning to perform your instrument - the computer - during this process?

A1 I am learning what the system is capable of doing, but it isn't a musical instrument: I do not shape sounds and material. I off-load those moment-to-moment decisions to the computational process. I learn how they sound, in relation to what you do, and develop them (between sessions. I don't alter the code live in performance) in relation to what I hear and what I would like to hear.

A2 In other words, this performance is a trio not a duo: you, me and your computer.

A1 In a sense, yes. But it is an unequal trio. The computer and myself are partners in the creative process. It is a co-creative system where I have overall executive control. And while the sounds we contribute are more varied than yours, I think your contribution carries more musical weight. What I mean is that you mostly lead, even when you take inspiration from the system. You also provide the context within which we hear the electronic sounds and judge whether the dialogue works or not.

A2 An important question is what our listeners get from our duo improvisations performances in terms of emotional and aesthetical engagement. How do they respond to the extra-human aspect of it? As such research in the context of ensemble improvisation performed by humans exists, it will be fascinating to compare. It begs for an audience research component, don't you think?

A1 It is a good question. And I wonder what is the effect of my presence at the computer on stage. We can tell the audience that my computer is not a musical instrument but what they see are two humans, each working on a machine. So your understanding of the extra-human aspect could be different to the audience's (and to mine). So if we want to study what is taking place we need to consider very carefully how we present the work and what we are asking.

A2 I am also wondering where is Ben-Tal in this improvisation? In what ways would it be different if another composer created such an artificial improviser?

A1 I embed my aesthetics and compositional ideas into the code. The code is an extension of me as a composer and it enacts what I want to hear in conjunction with your performance. The system tries to listen to you and the way it 'listens' reflects my priorities as a composer. Heavily constrained by the limitations of MIR techniques and my ability as a programmer. More significantly, the responses the system generates use the data that comes out of this listening process, but recombines this data in ways that reflect my ideas as a composer. Finally, during the performance I adjust aspects of the system in real-time based on what I would like to hear. The system is modular - I have about a dozen independent processes that I can turn on or off. Each one takes something from you and uses it in particular ways. I listen to the current music and imagine what adding or removing

a module would do. Or I imagine the outcome of any adjustments to the parameters I have control over. For example, for one of the modules I can change the register (octave transposition) of the resulting sounds. I can decide to place those sounds in the same register I hear from you or shift them to a different one. Of course, I can't predict whether you will continue in a similar manner or decide to change register in response.

A2 Can you explain what you mean that the code reflects your ideas as a composer?

A1 My introduction to computer music was through algorithmic composition - finding ways of enacting the sounds, materials, processes and time-structures I was after through lines of code. This is an iterative process: run code, listen to the result, change the code to improve the result. After a few iterations hearing fatigue makes it difficult to listen critically and evaluate whether the outcome is better. Another challenge is that for every sound you want to create you have to specify many parameters. Depending what technique you use for the synthesis you might need dozens of parameters for each sound. My solution to both these challenges was to include constrained randomness in the generation. I would directly specify some parameters - parameters that are important in this particular moment. Others would be chosen randomly each time I evaluate the code, though the choice would be constrained. The result would be slightly different each time, allowing me to listen critically. And I would adjust the parameters until the result would almost always be satisfactory. This approach is still there in this AI system. The responses it generates are musically appropriate and predictable on a statistical level. The precise events are not predictable but the overall effect is. And the system is able to both follow your lead and surprise you at times because it is computationally stable, the machine listening embedded in it makes it responsive, and it encodes musical ideas that you find engaging.

Coda

"Developing computer systems that collaborate in musical improvisation is a complex and challenging task due to the spontaneous and immediate nature of the musical dialogue required."¹⁷ In this paper, we describe the first stages of a successful development of such a system. The system is specific and tailor-made; it was designed to work with a specific human performer on a specific instrument (the piano). It also encodes specific compositional ideas: encompassing the type of material generated, the sonic properties, as well as the nature of the dialogue with the human improviser. Improvisation is a powerful test-bed for music AI because it is such a complex and challenging task. But also because improvisation encompasses many of the key components of musicality: listening, discerning, imagining.

Our system learns from the human performer, but does not employ the latest machine learning techniques with their reliance on extensive data and high-end computational power. Instead the system deploys older approaches to AI (e.g. rule-based systems; Markov-based models) in inventive ways. The learning aspect in the system is based on machine listening coupled with musically sophisticated uses of the data extracted - live and in real-time from

¹⁷ Gifford, T., Knotts, S., McCormack, J., Kalonaris, S., Yee-King, M., & d'Inverno, M. (2018). Computational systems for music improvisation. *Digital Creativity*, 29(1), 19-36

the performer - to generate novel material. Material that is novel but relevant to the musical context. In other words, while we see impressive improvements in the capabilities of AI systems through the combination of large data-sets and ever more powerful computing power, other approaches have not been exhaustively explored. And these approaches have the added bonus of being explainable and transparent. Unlike black-box models, Ben-Tal is able to modify the system to implement his musical ideas.

This AI system analyses data and then generates new data based on that analysis. However, while most recent music machine learning research invest the most effort - in terms of time, computing power, data acquisition, and innovation - into the learning (i.e. analysis) part, Ben-Tal's approach is the opposite. He invests time, creativity and thought in the music generation side of the pipeline. Prioritising the learning part is based on an implicit assumption that constructing better models will yield better outputs. However, when the definition of a successful model is not clear - as is the case for creative AI - this might not be the best route.¹⁸

Our research benefited from the integrated and collaborative development process. The AI system was (and still is) constantly tested on the basis of music-making. We trial the system during regular rehearsals where both authors hear the combined result (we also recorded some of the rehearsals allowing for repeated listening) and discuss them. We also invited friends and colleagues to some rehearsals and performed in front of audiences in concert situations. In addition, the fact that Dolan came into this process with little knowledge of AI or music technology required Ben-Tal to find ways of explaining what the system does and how it does it in musical and non-technical terms. We hope to see more research in AI music creativity that integrates the musical tasks and musical questions as part of the research methodology in similar ways.

Finally, our system benefitted from a slow development process with gradual improvements and refinement. As Professor Sageev Oore - Canada CIFAR AI Chair, and visiting research scientist at Google Brain - highlighted¹⁹ successful music making tools - such as the violin or the piano - were developed slowly over time. Both performers and composers contributed directly and indirectly to this development process. Therefore, It is unrealistic to expect music AI technology to emerge fully formed from a lab. We would hope to see a more sustained effort to involve musicians actively in the development process and allow time for mastering the creative potential of music AI applications.

Two important aspects of improvisation are missing from the AI at the moment. One is the ability to consider larger musical context - what we might describe as narrative or structure. This relates to both the listening and the generation parts of the system. We want to expand the machine listening beyond the local context and try to extract musical trajectories and identify sections and transitions. This should involve integration of different musical parameters as harmony, melody, tempo and texture are often interleaved in defining these larger scale aspects. For the generative side of the system, a good improviser should be able

¹⁸ Yin, Z., Reuben, F., Stepney, S. et al. Deep learning's shallow gains: a comparative evaluation of algorithms for automatic music generation. *Mach Learn* (2023).

<https://doi.org/10.1007/s10994-023-06309-w>

¹⁹ Personal communication with Ben-Tal, April 30th 2022.

to generate a plan for the longer-term while updating it constantly based on the developing context.

The second aspect still missing - and which also relates to both the listening and generating sides - is nuance. A more nuanced listening to the performer will aim to identify expressive intent achieved through means such as articulation, prosody or rubato. But, in the absence of a score to provide a framework, there is no obvious way of identifying these computationally. Applying such performance nuance to the generated material is a less daunting task technically, but the question of what would be the machine equivalent of expressive intent remains open. Though it could potentially be linked to the previous aspect of musical trajectories or structure

Finally, we want to add the dimension of audiences' evaluation and responses to the musical dialogue on stage. To what extent does the audience share our perception of musical dialogue between human and AI? Is this perception germane to the audiences' engagement with the music? If we want AI technology to part of the future of music we should consider its development in relation to both musicians as well as the audiences.