Using raw audio neural network systems to define musical creativity

Mia Windsor
Department of Music (Masters student)
University of Leeds
miarose.windsor@gmail.com

Abstract

This paper will use the hacker-duo Dadabots (who generate raw audio using SampleRNN) and OpenAI's *Jukebox* project (who generate raw audio using a hierarchical vq-vae transformer) as case studies to assess whether machines are capable of musical creativity, how they are capable of musical creativity, and whether this helps to define what musical creativity is. It will also discuss how these systems can be useful for human creative processes. The findings from evaluating Dadabots' and OpenAI's work will firstly demonstrate that our assumptions about musical creativity in both humans and machines revolve too strongly around symbolic models. Secondly, the findings will suggest that what Boden describes as 'transformational creativity' can take place through unexpected machine consequences [1].

1 Introduction

The hacker duo Dadabots use a modified SampleRNN (recurrent neural network) architecture to generate raw audio in the time domain [2]. The SampleRNN is trained to predict what audio sample comes next, one sample at a time, in a piece of audio divided at a sample rate of 16,000, 32,000 or 44,100 samples per second. During the training process, 'the gradient from this prediction loss backpropagates through time' or rather, it updates the choices it should have made to provide a correct prediction [3]. Dadabots use SampleRNN to generate audio in 'sub-genres with subtle stylistic distinctions such as black metal, math rock, and skate punk' [4]. Their work is regarded as important because of the inability of MIDI and symbolic models to create the 'nuanced spectral characteristics' found within the timbre and production of such genres. It is also important because of its ability to arguably transform beyond its source material through 'faults' or glitches in the neural network [5]. OpenAI's *Jukebox* generates songs using a hierarchical vq-vae transformer raw audio neural net to 'produce songs from highly diverse genres of music like rock, hip-hop, and jazz'. They use their models to capture melody, rhythm, long-range composition, and timbres for a wide variety of instruments, as well as the styles and voices of singers. Their models have been trained on over 700 bands [6]. Dadabots acknowledge that *Jukebox*'s hierarchical attention makes it better at song structure than SampleRNN's hierarchical LSTM [long short-term memory] [7]. I believe both of these case studies to be key in furthering discussion of the potential for AI Music Creativity because both are able to generate arguably creative and aesthetically interesting music. I think Dadabots' work has been especially important due their strong emphasis on the aesthetic impact, transformational capabilities and possible uses for their software.

This paper will analyse Dadabots' and OpenAI *Jukebox*'s outputs as case studies in order to challenge Margaret Boden's definition of creativity and David Cope's definition of musical creativity [8]. Section 3 will discuss both case studies' capabilities in combinatorial creativity (as defined by Boden) and will argue that Cope's definition of musical creativity is outdated because of its basis in MIDI and symbolic models. This section will also challenge Merker's argument that music is a Humboldt system; a notion that is tied to Cope's argument [9]. Section 3.2 will acknowledge both system's abilities (in *Jukebox*'s case) or lack thereof (in Dadabots' SampleRNN's

case) to allude to other works and will argue that though allusion can be important, there are other ways of getting out of search spaces [10]. Sections 4.2 and 4.3 will review the outputs mentioned in OpenAI's 2020 paper and Dadabots' 2018 paper and will suggest that Dadabots' SampleRNN's and OpenAI *Jukebox*'s outputs are capable of transformational creativity as defined by Boden. Section 4.4 will argue against Boden's claim that the conceptual space in question must be 'fully explored' in order for transformation to occur, as this is not the case with unexpected machine consequences [11]. Section 4.5 will discuss other methods for transformational creativity in machine-made music. Finally, Section 5 will highlight how raw audio neural network's creative capabilities can be useful for expanding possibilities in music.

2 Existing definitions of musical creativity

When defining creativity in general, the broad consensus is that creativity is the capacity to produce things that are original and valuable [12]. Although these words mean different things, there is certainly a link between value and originality; for example, the level of originality of a work of art can help shape the aesthetic value. Boden hypothesises that there are three types of creativity: combinatorial (when novelty arises out of unfamiliar combinations of ideas); exploratory (creating ideas through existing conventions); and transformational or 'radical' (altering the conceptual space itself, such that new ideas are generated that do not fit into a previous style or convention) [13]. Boden also argues that the artist or inventor must fully explore the conceptual space in question before transformation can occur [14]. Boden's descriptors have been used in this paper as they provide very clear distinctions for categorising the work of Dadabots' and OpenAI's *Jukebox* project. Boden's description of transformational creativity will be built on in section 4.3.

Cope's definition of musical creativity is the only definition I could find that focuses on music specifically and is non-human-centric. It will be used to ground the discussion whilst also being built on. Cope's definition of musical creativity is process-based; he defines it as 'The initialisation of connections between two or more multifaceted things, ideas, or phenomena hitherto not otherwise considered actively connected [15]'. The multifaceted idea is significant to acknowledge because music functions on the horizontal and vertical planes [16]. It is important to establish that Cope's book *Computer Models For Musical Creativity* focuses on processes for musical creativity in machines and uses this to define musical creativity. His arguments are based entirely on his *Experiments In Musical Intelligence* and *Emily Howell* programs. In his book, it is arguable that *Emily Howell* is capable of musical creativity through the ability to synthesise the work of others, to juxtapose this with allusions to different works, to take influence from wider contexts, and to extend rules using analogies [17].

3 Combinatorial creativity

It could be argued that Dadabots' and Open AI *Jukebox'*s raw audio neural networks are simply examples of combinatorial creativity on a micro scale because they are a re-combination of tiny samples of audio. Cope states that 'The secret of successful creativity lies not in the invention of new alphabet letters or musical pitches, but in the elegance of the combination and recombination of existing letters and pitches [18].' This is reminiscent of Merker's argument that music is a Humboldt system, or rather, a system that can generate infinite diversity by finite means, where the phenomena do not blend their properties [19]. Though Cope's argument is true for some MIDI and symbolic models, it is important to note that raw audio neural networks are instead working with a much more varied palette of sound, because working with raw audio also allows the emulation of timbral and production techniques [20]. Merker's argument is technically still valid in the case of raw-audio neural networks as there is still a finite number of samples from the training data. Though the samples also do not blend by averaging their properties, my perception of the music suggests otherwise: in my listening experience, the patterns are so small and intricate that they do appear to blend, even if this is not what is actually happening. This will be further discussed in Section 4. It

is also noteworthy that methods do exist for technically blending sounds together, such as through convolution. Convolution is where the frequency spectra of two audio sources are multiplied resulting in the frequencies that are shared between the two sources being accentuated, and the frequencies that are not shared being attenuated, thus, literally blending audio sources. This therefore suggests Merker's argument to be redundant. Merker's argument is also a direct attempt at disregarding Hanslick and Varese's definitions of music, which both include 'noisy' sounds. Merker is opposed to these definitions because they do not distinguish music from speech [21]. Though Hanslick and Varese's definitions could be argued to be extreme, it is significant to acknowledge that timbre and production, and thus, more noisy and complicated timbres, are often paramount in popular music (and a lot of contemporary music) and are often too complicated to be worth representing in MIDI or symbolic form (an obvious 'Humboldt system'). The assumption that a piece of music can be represented easily in terms of melody, harmony, rhythm, texture, dynamics and simple instrumentation is extremely outdated and mostly only applicable to classical music. This therefore suggest both Cope and Merker's arguments to be outdated. It is also important to acknowledge that both Dadabots SampleRNN and OpenAI's *Jukebox* systems go beyond human capability with their timbre and production emulation. Working at the sample level to *almost* exactly re-create particular sounds such as an individual's voice is a technique that is inaccessible to humans because we are unable to work with and perceive such small sections of audio. Any attempts at such have used logical processes, have required extraordinary effort, and have not succeeded in producing or reproducing anything familiar [22]. This suggests these particular combinations of sound to be beyond anything a human could physically put together.

3.2 Allusion

In their papers, Dadabots establish that their neural network was 'trained on whole albums from a single artist' to 'synthesise music within the limited aesthetic space of the album's sound' [23]. If their output *is* regarded as creative, this contests Cope's view that it is 'difficult to imagine a program exhibiting creative-type behaviour not having access to the diversity made available by multicomposer databases', because 'creativity relies in part on the juxtaposition of allusions to the work of others' [24]. OpenAI's *Jukebox* was trained on over 700 bands and is therefore always subtly alluding to an array of works making it successful in this regard. *Jukebox* can merge two musical genres by adding an artist's generated voice, riff, or drum pattern into a track based on a different artists style. This demonstrates clearer allusion. *Jukebox* is also able to blend voices or create vocal 'duets' between different artists' generated voices [25]. This attempt is much closer to the goal of distinctly alluding to other works, though it seems that often only two works are merged to prevent the audio from getting messy. Dadabots' SampleRNN has been less successful in its attempts at allusion as demonstrated by its attempt to make something coherent out of the Beatles' greatest hits [26]. It is again important to note that the building blocks are smaller in the case of both Dadabots SampleRNN and OpenAI's *Jukebox*, and their 'databases' are likely to have as many components as Cope's 'multicomposer databases'. It could also be argued that allusion is just one method of getting out of a confined search space and that there are other methods that are equally as important. Some of these will be discussed in the following section. Perhaps then, allusion is a useful, yet not completely necessary tool for achieving creativity.

4 Transformational creativity

Cope states that 'In order for computer programs to create, they must themselves develop and extend rules, and not simply follow instructions provided by programmers' [27]. This point is interesting when it comes to Dadabots. Their initial intention with the neural network was to create 'a realistic recreation of the original data'[28]. However, judging by the music itself, and their observations, it appears that the faults of the neural network, as can be the case in the human neural network, have allowed the source material to transform, perhaps beyond combinatorial creativity. In this case, the transformation results in a specific, visible and clearly inhuman aesthetic. As they state themselves:

> Solo vocalists become a lush choir of ghostly voices, rock bands become crunchy cubist-jazz, and cross-breeds of multiple recordings become a surrealist chimera of sound. Pioneering artists can exploit this, just as they exploit vintage sound production (tube warmth, tape-hiss, vinyl distortion, etc). [29].

Though Dadabots do highlight the aesthetic merit of the SampleRNN's imperfections, it is interesting that they only compare these to production techniques rather than alterations of the composition itself. It is clear that Dadabots' SampleRNN, through its partial 'failure', is able to make changes to the composition as well as the production, such as the 'abrupt sectional changes, odd meters, and long rests' changing a three-song album by rock band Room For a Ghost into 'math rock' [30]. These changes could arguably be described as transformations of the training material. Though OpenAI do not highlight the aesthetic merit of the imperfections of *Jukebox*, similar transformational qualities are also present in their outputs. The following section will review the tracks created by *Jukebox* discussed in their 2020 paper and the albums created by Dadabots discussed in their 2018 paper [31]. OpenAI state that 'because everyone experiences music differently, it is generally tricky and not very meaningful to evaluate samples by the mean opinion score or FID-like metrics'. They proceed to evaluate tracks generated by Jukebox manually, taking into consideration coherence, musicality, diversity, and novelty [32]. This section will continue discussing the outputs of both OpenAI's *Jukebox* and Dadabots' SampleRNN. It will loosely take into account the criteria set out by OpenAI and will attempt to shed light on transformational creativity potential and its impact on me as a listener.

4.2 ***Jukebox*** **Samples**

The curated samples by OpenAI's *Jukebox* are certainly clearer and more accurate than the Dadabots albums. Many of the samples are 'successful' in their attempts to re-create the sounds of particular artists and styles such as *Heavy Metal, In The Style Of Rage,* both *Country, in the style of Alan Jackso*n tracks, *Pop, in the style of Céline Dion,* and all of the song completions. I was especially impressed by *Heavy Metal, In The Style Of Rage*. The track was incredibly convincing and complex, with clear textural contrasts in the structure, well developed riffs and parallel harmonies cutting through clearly. Interestingly, the track started off sounding like metal then gradually moved into sounding more like emotion-heavy classic rock. Though there are certainly pointers towards classic rock in Rage's original sound, it was an odd transition. This track was also successful in its variety in vocal timbre with some interesting backing vocal interjections throughout the track which were distinctly different from the lead vocal.

Though these 'accurate' re-creations are impressive, I was more a fan of *Jukebox*'s attempts to create songs 'by' artists whose original recordings had more low-fi production as this seemed to garble the output in interesting ways. This was the case with *Classic Pop, in the style of Frank Sinatra, Rock, In The Style of Elvis Presley* and *Jazz, in the style of Ella Fitzgerald*. The murky sounding string stabs at the start of *Pop, in the style of Frank Sinatra* are both distant and dissonant to a point where you can't actually tell what chord is being played. I enjoyed this ambiguity. There's also some interesting timbral effects throughout such as crackling or laughing sounds occasionally penetrating through. *Jazz, in the style of Ella Fitzgerald* pretty much turns into an experimental noise track from 3:12 onwards with noises that sound like a release of air or steam, occasional gentle double bass from the original track, and pitched percussive sounds that sound like they were created using an FM synthesiser. The track then builds into a brief rhythmically unaware garbled drum idea followed by applause. I liked this very short section a lot and would genuinely consider sampling it. I would argue it to be a successful transformation of the original material as some congruity is maintained from the distinct double bass sound while being morphed into something unexpected.

4.3 Dadabots albums

4.3.1 *Codanity of Timeness*

Trained on the black metal album *Diotima* by Krallice.

Though the original album opts for more hi-fi production, the Dadabots album results in a more low-fi sound that is more typical of the black metal genre. Much of the Dadabots album is resemblant of the final track *Dust and Light* on the original album which is more atmospheric in its nature than the other tracks due to a looser harmonic and rhythmic structure. The harmony in the Dadabots album blends together in quite a sophisticated way. Though the structure is not as distinct as the melody and harmony of the original power chords, there are still subtle pitches that occasionally cut through. This album provides an especially clear example of samples that are so small and intricate appearing to blend into each other in an atmospheric wash of sound, suggesting then that, as opposed to Merker's argument, perceptual blending *is* possible in music. I would also argue that these 'blends' allow opportunity for transformation.

4.3.2 *Inorganimate*

Trained on the math metal album *Nothing* by Meshugah.

This album is a human-curated combination of different epochs (checkpoints in the training process). Dadabots state that:

> The earlier epochs had weird texture and percussion effects. The middle epochs came up with tempos and vocal styles not used in the album. The later epochs better replicated the band's palette but arhythmically and awkwardly (though the effect is humorous) [33].

In the Dadabots album, *Disenchameleons* and the end of *Clost* likely came from the earlier epochs due to their chaotic sound. The unfamiliar vocal styles are certainly present in *Dehumanizational*. The most interesting aspect of this album to me is its use of quiet sounds, that do not predominate in the original. These sounds could have originated from many different places in the original such as the fade-outs of tracks, the end resonating note in *Straws Pulled at Random*, or even just in one of the quieter parts of the album. In the Dadabots album, *Dehumanizational* features an eerie quiet buzzing sound that comes in a few times, occasionally overpowering the material and then subsiding, as if the track is fighting against it taking over (which it eventually does). This makes the track especially unsettling. I found this one particularly interesting because it had such an unexpected effect on me that was not present in my experience of the original album.

4.3.3 *Megaturing*

Trained on experimental rock album *Mirrored* by Battles.

In this album, the human-curated part is extended by 'introducing a new audio layering technique intended to create a stereo image from monophonically generated samples' [34]. The tracks *Dominal*, *Makines* and *Electic* in the Dadabots album are especially noteworthy.

I would argue that this album takes the characteristics of math rock and brings it to an odd boundary where you can't always tell if it's become rhythmically more complex than the original or just a confused mess. In either case, it sounds alien. *Dominal* exploits the main sample found in *Ddiamondd* but in a more chaotic manner as if the sample is smeared across the time domain, again giving a suggestion of multiple sounds 'blending' together. The faster rhythm and distortion is actually more reminiscent of a metal album, which shows a clear transformation. *Makines* comes

across as an extremely inhuman variation on *Tij* from the original album. *Makines* is faster with any regularity present in *Tij* removed from the rhythm and tempo, making it even more extreme. Finally, *Electic*, is interesting in terms of tempo. It acknowledges the riff in *Tonto* (that gradually slows down) but instead presents it with irregular fluctuations in speed. This creates an uneasy stuttering effect. This album is a prime example of alteration of the composition rather than the production. I would argue that this alteration is more than just simple variation as it again, is incredibly intricate, unsettling and arguably, transformational.

4.3.4 *Bot Prownies*

Trained on skate punk album *Punk In Drublic* by NOFX.

This album arguably just sounds like bad skate punk. However, it is hilarious and the *track Loose Home* (as an example) sounds like a genuine parody of lead singer Fat Mike's vocal style. The fact that a SampleRNN can turn raw audio into something comedic is impressive and Dadabots have stated themselves that they are usually 'motivated by comedy' [35]. Interestingly, this album functions aesthetically in a similar way to mash-ups. Ragnhild and Harkins point out that the reason mash-ups are often humorous is because of the contextual incongruity of recognisable samples and the musical congruity between the mashed tracks [36]. The incongruity of recognisable samples is relevant here: the vocal style in this album is recognisable to anyone familiar with the stereotypical Californian punk vocal style; however, this voice is put in the new context of garbled vocal sounds rather than actual words [37]. This is what creates the incongruity and thus, the humorous effect, in turn transforming the purpose of the material. The musical congruity between the tracks is less relevant as all the samples come from the same album. More generally, the retention of the band's sound in these albums is aesthetically similar to plunderphonics in that they both act as referential sound collages of existing work with the recognisability coming from timbre [38]. The work of Dadabots just acts on a much smaller scale. These collages, as done in plunderphonic and mashup works, allow the context of the original work to be shifted which is arguably transformative.

4.4 Errors as a transformational tool

Though I would encourage you to have a listen yourself and make your own judgement, these examples do suggest to me that raw audio neural networks are capable of what Boden describes as 'transformational creativity' [39]. Through their imperfections, these systems have been able to go beyond their training material to create something that appears to be transformational. However, unlike Boden's 'requirements', these transformations occur through errors, rather than fully exploring their 'conceptual spaces' [40]. Novitz argues that Boden's perspective on conceptual space exploration is problematic because in some cases, close acquaintance with conceptual spaces may actually inhibit creativity, especially when it comes to human creativity [41]. Shedding more light on this from a computational perspective, Holly Herndon states that when training her AI *Spawn* to sing, she avoids using classical music canon as training data, which one would assume avoids exploration of part of the conceptual space, that space being existing vocal technique [42]. However, *Spawn's* interpretation and timbral abilities are unique in comparison to a classically trained singer, therefore implying that this software produces creative originality owing to its isolation. This, therefore, contradicts Boden's view that the conceptual space must be fully explored. The idea of conceptual spaces being transformed through errors is reminiscent of Leijnen's view that systems may exist which transform the search space (the data available to the system) by generating and eliminating constraints, acknowledging 'bugs, errors and random numbers' as a means of getting out of a designed search space [43]. Although Leijnen states that 'bugs, errors and random numbers' are ultimately unsuccessful because they 'do not specify where to get to, or how to get there', the changes demonstrated by these raw audio neural networks suggest that transforming the search space with bugs, errors and random numbers is possible in a meaningful way [44].

It is relevant that many examples of musical creativity in the age of technology have come about through humans recognising creative potential in the sound of something going 'wrong' or 'unexpected' and exploiting this. Both case studies do exactly this, though the recognition of the error is of course not conscious. Examples of other occurrences of transformational creativity due to unexpected behaviour or errors in technology include Grandmaster Flash's discovery of "scratching" (using records as sources for percussive expressive sounds), or the exploitation of the Roland TB-303 by electronic dance music artists creating the acid house genre, taking advantage of the TB-303's 'squelchy tone' which had been deemed unsuccessful in its original purpose of simulating bass guitars [45]. Another relevant example is the emergence of the glitch as an aesthetic or conceptual element in music through damaging CDs or collapsing software processing [46]. A lot of the sounds present in both Dadabots' and *Jukebox*'s outputs could be described as glitches as they are instances of hardware or software not working in the way that is expected, producing a sound that is so inhuman that it sounds very distinct to the human listener. Dadabots' SampleRNN and OpenAI's *Jukebox* project relate especially to what Sanglid describes as 'oceanic glitch', where glitches are worked through compositionally to create relatively coherent music. The music of Oval is a good example of 'oceanic glitch': typically they take CDs containing existing music, damage them, then loop samples of them to create ambient 'dense, flitting textures', resulting in an aesthetic that shares similarities with the music of Dadabots and OpenAI's *Jukebox* due to their intricate glitches [47]. It is very important to note that these raw audio neural networks are responsible for every step of this transformational process. Unlike Oval, it is not the humans working the glitches into the composition and rather the machines themselves. Though human curation is sometimes useful for larger-form structures, this is not always necessary. This therefore suggests raw audio neural networks to be independently capable of transformational creativity.

4.5 Other methods of transformational creativity by machines

Sanglid describes genetic mutation as a glitch that can sometimes turn out to be beneficial for the species, stating that 'evolution depends on errors and imperfections' [48]. It seems important to acknowledge that such genetic glitches can be emulated computationally when developing musical material, through the use of genetic algorithms (GAs). GAs may be useful for transforming music beyond its source material as an addition to other AI and generative systems, especially those that use MIDI and symbolic models where there is less complexity, and thus, less potential for glitches to occur on their own. Though GAs are unable to learn, they can develop beyond a source material and the listener's expectations by breeding and mutating musical ideas over multiple generations. Entities that are selected to crossover or 'breed' are determined by a fitness function that can be based on rules, randomness, or human input [49]. If the fitness function and mutation functions are random, the algorithm is just a form of controlled randomness, yet it is important to note how sophisticated these algorithms can be. In the case of a random fitness and mutation function, new material may not just be a few slightly different pitches and rhythms but may have been recombined and altered to transform into something regarded as completely novel. This, therefore, suggests that genetic algorithms are an effective way to achieve transformational creativity.

In addition to this, it is clear that the association networks in Cope's *Emily Howell* system are an alternative and very sophisticated method for transformational creativity, where the system can recognise analogous musical patterns and feeds itself through recursion. It can also make analogies to works that are not in the primary database for the purpose of allusion or providing more general context [50]. Although Cope's system is arguably a sophisticated method for transformational creativity in machine composition, it may not work beyond the constraints of classical music. There are two reasons for this. Firstly, because other styles of music follow much looser musical patterns which are more difficult to analyse. Secondly, because as discussed previously, timbre and production techniques can be paramount in certain styles of music, which can result in an inability to communicate the music successfully in MIDI or symbolic form. In either of these cases, naturally occurring glitches or genetic algorithms may be a better approach for applying transformational creativity in a sophisticated way to different styles of music.

5 Why are these systems useful

It is important to discuss the point made by Dadabots that their SampleRNN system can be exploited by pioneering artists. This is important because it emphasises the fact that these systems can benefit humans' creative processes rather than passively generating music with little input from the user. Arguably the most creative exploitations of Dadabots' SampleRNN, thus far, have been by solo artists. Dadabots have collaborated with the UK beatboxer Reeps One, training the SampleRNN on his technique so that he can perform a duet with himself [51]. The alternations between the real vocal sounds and the machine interpretations create intriguing contrasts resulting in an engaging performance. Dadabots have also collaborated with the jazz bassist and YouTuber Adam Neely to create a live stream of an infinite bass solo that was initially fed by two hours of Neely's playing [52]. Neely spends a lot of time in his video about the collaboration taking samples from the live stream and playing alongside the samples to create new tracks; he also actively encourages the viewers to sample it themselves [53]. This is especially significant because it shows that the use of raw audio neural networks can be an effective creative tool for human composition. More generally, it seems that combinatorial and transformational creativity in machines may help us develop our own, other artists', or other machines' material in ways we cannot predict due to our preconceived expectations, or other machines' limitations. I think Dadabots' attempts at directly reaching out to other artists has been incredibly productive in shedding new light on the potential of their SampleRNN system and I would encourage OpenAI to do more of the same with the *Jukebox* project.

6 Conclusions

In conclusion, this paper has firstly argued that raw audio neural networks are at least capable of combinatorial creativity due to their ability to combine small samples of audio, in a unique way that is difficult for humans to achieve. This also suggests that traditional assumptions regarding musical creativity (especially those demonstrated by Cope) revolve too strongly around symbolic representations of composing and should also consider timbre and production. The view demonstrated by Merker that music is a Humboldt system has also been problematised by the existence of raw audio neural networks due to their suggested ability to perceptually blend sounds. Secondly, this paper has suggested that raw audio neural networks are capable of transformational creativity through subtle glitches and imperfections, which therefore implies that conceptual spaces do not have to be fully explored for transformation to occur. This paper has also suggested some alternative methods for transformational creativity in machines. Finally, this paper has evidenced how the combinatorial and transformational machine possibilities discussed can be useful for human musicians or for diversifying the outputs of other machine systems.

References

[1] Margaret Boden, *The Creative Mind: Myths and Mechanisms* (London: Routledge, 2004), p. 3

[2] CJ Carr and Zack Zukowski, 'Generating Black Metal and Math Rock: Beyond Bach, Beethoven, and Beatles', *31st Conference on Neural Information Processing Systems* (2017) <https://arxiv.org/abs/1811.06639>, p. 1

[3] CJ Carr and Zack Zukowski, 'Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands', *Proceedings of the 6th International Workshop on Musical Metacreation* (2018) <https://arxiv.org/abs/1811.06633>, p. 2

[4] Ibid, p. 1

[5] Ibid, p. 1

[6] Prafulla Dhariwal and others, 'Jukebox: A Generative Model for Music' (2020) <https://arxiv.org/abs/2005.00341> [11 January 2021]

[7] DADABOTS, 'lofi classic metal ai radio – riffs for false nostalgia 24/7', YouTube (2020) <https://www.youtube.com/watch?v=J1NV6CUJl18> [10 January 2021]

[8] Boden, *The Creative Mind* (2004); David Cope, Computer Models of Musical Creativity (Cambridge Massachusetts: MIT Press, 2005), p. 11

[9] Bjorn Merker, 'Music: The Missing Humboldt System', *Musicae Scientiae,* 6.1 (2002) 3-21 (p. 4)

[10] Cope, p. 258; Cope, p. 125

[11] Boden, *The Creative Mind* (2004), p. 3

[12] Berys Gaut, 'The Philosophy of Creativity', *Philosophy Compass*, 5.12 (2010), 1034-1046 (p. 1039)

[13] Boden, *The Creative Mind* (2004)

[14] Margaret Boden, *The Creative Mind: Myths and Mechanisms* (Reading: Cardinal, 1992), p. 46, quoted in David Novitz, 'Creativity and constraint', *Australasian Journal of Philosophy*, 77.1 (2006) 67-82 (p. 71); Margaret Boden, 'What is Creativity', in *Dimensions of Creativity*, ed. by Margaret Boden (Cambridge, Massachusetts: MIT Press, 1996), pp. 75-118 (p. 80), quoted in Novitz, p. 71

[15] Cope

[16] Ibid

[17] Ibid

[18] Ibid, p. 89

[19] Merker, p. 4

[20] Carr and Zukowski, 'Beyond Bach, Beethoven, and Beatles', p. 1

[21] Merker, p. 3; Eduard Hanslick, *Vom Musikalilisch-Schönen. Ein Beitrag zur Revision der Aesthetik der Konkunst* (Leipzig, 1854); Edgard Varèse, 'Organized sound for the sound film', *Commonweal*, 33 (1940) pp 204-205

[22] Curtis Roads, *Microsound* (Cambridge, Massachusetts: MIT Press, 2001), p. 31

[23] Carr and Zukowski, 'Beyond Bach, Beethoven, and Beatles*',* p. 2; Carr and Zukowski, 'Generating Albums with SampleRNN', p. 2

[24] Cope, p. 258; Cope, p. 125

[25] Dhariwal and others

[26] Dadabots, *Deep the Beatles!* (2017)

[27] Cope, p. 271

[28] Carr and Zukowski, 'Beyond Bach, Beethoven, and Beatles*',* p. 2

[29] Ibid

[30] Ibid; Neither the original nor the SampleRNN generated album now exist.

[31] Dhariwal and others; Carr and Zukowski, 'Generating Albums with SampleRNN'

[32] Dhariwal and others

[33] Carr and Zukowski, 'Generating Albums with SampleRNN'

[34] Ibid, p. 3

[35] Adam Neely, 'Making a NEVER-ENDING Neural Net Bass Solo (ft. Dadabots)', *YouTube* (2020)

<https://www.youtube.com/watch?v=2xMhRwxXJTc&t=840s> [10 January 2021]

[36] Ragnhild Brøvig-Hanssen and Paul Harkins, 'Contextual incongruity and musical congruity: the aesthetics and humour of mash-ups', *Popular Music*, 3.1 (2012), 87-104 <http://dx.doi.org/10.1017/S026114301100047X>

[37] Dan Nosowitz, 'I Made A Linguistics Professor Listen to a Blink-182 Song and Analyze the Accent', *Atlas Obscura* (2015) <https://www.atlasobscura.com/articles/i-made-a-linguistics-professor-listen-to-a-blink-182-song-and-analyze-the-accent> [3 April 2021]

[38] Kevin Holm-Hudson, 'Quotation and Context: Sampling and John Oswald's Plunderphonics', *Leonard Music Journal*, 7 (1997), 17-25 (p. 21)

[39] Boden, *The Creative Mind* (2004), p. 3

[40] Ibid

[41] Novitz, 67- 82 (p. 71)

[42] Sónar+D, 'Sónar+D Talks: Listening to the Voice of #AI, with Holly Herndon and Mat Dryhurst', *YouTube* (2019) <https://www.youtube.com/watch?v=k67IKMXzD7A> [23 October 2020]

[43] Stefan Leijnen, *Creativity & Constraint in Artificial Systems*, (unpublished doctoral thesis, Radboud University Nijmegen, 2014), p. 68

[44] Ibid

[45] Eliot Bates, 'Glitches, Bugs, and Hisses: The Degeneration of Musical Recording and the Contemporary Music Work', in *Bad Music: The Music We Love To Hate*, ed. by Christopher J. Washburne and Maiken Derno (New York: Routledge, 2004), pp. 212-225 (p. 219)

[46] Torben Sanglid, 'Glitch—The Beauty of Malfunction', in *Bad Music: The Music We Love To Hate*, ed. by Christopher J. Washburne and Maiken Derno (New York: Routledge, 2004), pp. 198-211 (p. 199)

[47] Ibid, p. 200; Kim Cascone, 'The Aesthetics of Failure: "Post-Digital" Tendencies in Contemporary Computer Music', *Computer Music Journal*, 24.4 (2000) 12-18, (p. 13)

[48] Sanglid, 198-211, (p. 207)

[49] John Al Biles, 'Evolutionary Computation for Musical Tasks', in Evolutionary Computer Music, ed. by Eduardo Reck Miranda and John Al Biles (London: Springer, 2007), pp. 28-51 (p. 42)

[50] Cope

[51] Swissbeatbox, 'Reeps One ft. A.I. 'Second Self' (We Speak Music | Episode 6 | Human and Machine)', *YouTube* (2019) <https://www.youtube.com/watch?v=q981cTdL0_Y&t=216s> [15 April 2021]

[52] DADABOTS, 'Infinite Bass Solo', *YouTube* (2020) <https://www.youtube.com/watch?v=RRwJzw4VvZA> [15 April 2021]

[53] Neely