Towards Sound Innovation Engines Using Pattern-Producing Networks and Audio Graphs*

Björn Þór Jónsson[1,2][0000−0003−1304−5913], Çağrı Erdem[2][0000−0003−2632−6829], Stefano Fasciani[3][0000−0001−5555−3225], and Kyrre Glette[1,2][0000−0003−3550−3225]

[1] RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway
https://www.uio.no/ritmo
[2] Department of Informatics, University of Oslo, Norway
[3] Department of Musicology, University of Oslo, Norway
{bthj,cagrie,stefanof,kyrrehg}@uio.no

Abstract. This study draws on the challenges that composers and sound designers face in creating and refining new tools to achieve their musical goals. Utilising evolutionary processes to promote diversity and foster serendipitous discoveries, we propose to automate the search through uncharted sonic spaces for sound discovery. We argue that such diversity promoting algorithms can bridge a technological gap between the theoretical realisation and practical accessibility of sounds. Specifically, in this paper we describe a system for generative sound synthesis using a combination of Quality Diversity (QD) algorithms and a supervised discriminative model, inspired by the Innovation Engine algorithm. The study explores different configurations of the generative system and investigates the interplay between the chosen sound synthesis approach and the discriminative model. The results indicate that a combination of Compositional Pattern Producing Network (CPPN) + Digital Signal Processing (DSP) graphs coupled with Multi-dimensional Archive of Phenotypic Elites (MAP-Elites) and a deep learning classifier can generate a substantial variety of synthetic sounds. The study concludes by presenting the generated sound objects through an online explorer and as rendered sound files. Furthermore, in the context of music composition, we present an experimental application that showcases the creative potential of our discovered sounds.

Keywords: Sound Synthesis · Quality Diversity Search · Innovation Engines.

1 Introduction

Either you know what sound you're looking for, or you don't know what sound you're looking for. In the latter case, inquiry, or prompting, is impossible. To

* Supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762.

discover new sounds, you must recognize them when you have found them. But if you can do that, you must have known them already. Transferring such a paraphrasing [30] of Meno's Paradox to the domain of novel sound design can be a way of establishing the usefulness of serendipitous sonic discoveries, where a new sound may not have been explicitly sought after but immediately recognised when heard. With all sound admissible as material for making music and all sounds theoretically possible with digital synthesis, there is still much more to explore considering the entirety of the sonic domain [38]. Composers and sound designers often need to create and refine new tools in order to achieve their musical goals. This endeavour may be hindered by a lack of technical expertise. Our proposed approach leverages evolutionary processes to generate novel sounds, thereby facilitating the creative journey and overcoming the technical barriers that may limit composers and sound designers in expanding their sonic repertoire.

We work towards an approach to automate navigation through previously unexplored sonic territories. As such, while entirely novel, the discovered sounds can be perceived as appealing and seemingly recognisable to the listener despite their unprecedented nature. Such investigations have been carried out interactively in the visual domain [34], demonstrating the usefulness of abandoning specific objectives, or at least switching goals as stepping stones are found while traversing paths to interesting discoveries. These findings provided a basis for proposing the Novelty Search algorithm [19] and later other variants, forming a family of Quality Diversity (QD) search algorithms [20,25,32,3]. These QD algorithms combine the open-endedness of Novelty Search with competition between solutions in their own "niche", resulting in diverse and high-performing (quality) solutions. Overall, QD algorithms serve as effective tools for illuminating high-quality solutions within a domain and are powerful search algorithms in their own right. This is due in part to their ability to exploit behavioral diversity and stepping stones during the search process, which can lead to discovering a variety of valuable solutions [7,29]. To drive automated exploration with such diversity-promoting algorithms, the Innovation Engine algorithm abstracts the process of human curiosity, replacing human judgement with a discriminative model that identifies interesting ideas [28,26]. Innovation Engines integrate two key components: Evolutionary Algorithms (EAs), such as those from the family of QD, capable of generating and gathering various novel outputs; and a model capable of distinguishing that novelty and evaluating its quality, such as Deep Neural Network (DNNs), creating niches and competition within them, thus providing selection pressure to quide QD search. The ultimate goal of this architecture is to continuously generate interesting and innovative creations in any given field.

Compositional Pattern Producing Networks (CPPNs) [35] are a foundation of the explorations leading to the Novelty Search and Innovation Engine algorithms. The networks abstract unfolding development in evolutionary processes, which build a phenotype over time. This is done by using any variety of canonical functions at each node, based on the idea that the order in which the networks compose functions can provide that abstraction. This can be compared with the

process of timbral development, where musical expression depends on changes and nuances over time. The use of patterns produced by CPPNs as sources of sound- and control signals for sound synthesis has been explored in a novelty seeking Interactive Evolutionary Computation (IEC) [37] configuration, which was inspired by previous work on the generation of visual artefacts [16]. The representation of temporal unfolding provided by CPPNs has been combined with the evolution of Digital Signal Processing (DSP) graphs during several iterations of investigation, detailed in [14]. This resulted in a distinct approach to sound synthesis, where any combination of the two graphs, depicted in figure 1, can be rendered at any duration, revealing the sub-patterns encoded by CPPNs over varying periods of time.

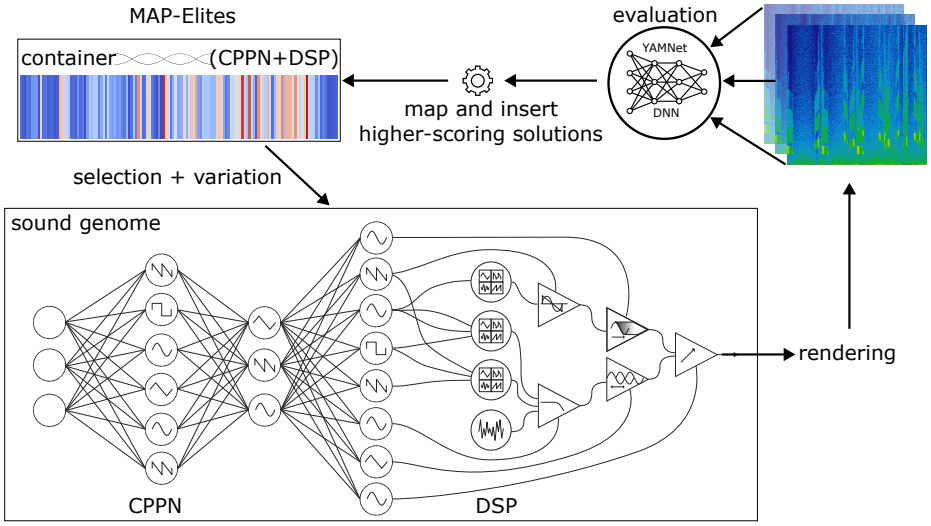


Fig. 1: The QD algorithm MAP-Elites uses the pre-trained YAMNet DNN classifier to define cells in a container and the performance of an evaluation candidate across those cells to determine placement and replacement in that archive. The genome of each evaluation candidate is rendered to a waveform, which is supplied to the classifier. Inputs to the CPPN are discussed in section 2.

Given the more diverse application of sonic artefacts as material in creative processes, we argue that there may be an even higher incentive to investigate the Innovation Engine algorithm's applicability in the sound domain. Furthermore, whereas humans can evaluate images in a split second, evaluation of sounds requires more time. There is a minimal duration threshold for perceiving salient features of sonic objects [24] and we typically perceive them holistically as meaningful units in the 0.5 to 5 seconds range [10]. Experiments with interactive evolution of sounds [16] revealed how fatigue can set quickly in when potentially listening to a long series of taxing sounds. This further limits the ability of hu-

mans to provide sufficient quantity of selection to have a significant effect on evolution.

Automating the discovery of new sounds is the goal of this study. We achieve this by applying the Innovation Engine algorithm to the sound synthesis approach developed in previous research on interactive novelty discovery. By using the proposed technique for sound synthesis, the system does not need to be trained beforehand as the evolutionary method starts from networks with no hidden nodes and progressively evolves primitive individuals by adding nodes and connections with the NeuroEvolution of Augmenting Topologies (NEAT) algorithm [36]. In our initial experiments, we use a signal from a pre-trained discriminative model to guide QD search, without human feedback in the evolutionary loop. Investigating this setup is intended to pave the way for further explorations of unbounded discovery of interesting sounds.

Our contributions include researching the application of a special type of Innovation Engine in the sound domain with a distinct approach to sound synthesis within an EA. Furthermore, we examine different configurations of our generative system and study how our sound synthesis method interacts with the discriminative model. We also offer a web-based interface to explore the outcomes of our evolutionary processes through our Innovation Engine setup. Lastly, we showcase audio artefacts rendered from the solutions discovered during the QD runs. Experimental results, in the form of historical data from evolution runs, elite maps and genomes from each point in time, and sounds rendered from those genomes at final iterations, along with the source code to replicate the results, are available in the dataset accompanying this article [15].

2 Approach and Experimental Setup

To start evaluating the applicability of the Innovation Engine algorithm in the domain of sounds, we combine a sound synthesis technique with a supervised discriminative model. The foundation of our sound-generating system relies on using the patterned outputs from CPPNs as the raw materials for sound and control signals. These signals can be utilised in their original form or further shaped through a DSP graph. Such a design choice enables the evolutionary state to begin from a blank slate, established with random initialization of the CPPN and DSP graph counterparts. This avoids dataset constraints that might limit the potential for discovery of novel sounds. The genome evolved by the evolutionary (QD) processes is composed of the CPPN and DSP networks and the evolvable connections between them. Details of this genome configuration are discussed and diagrammed in [14]. Figure 1 illustrates the data flow of our experimental setup and shows how the genome fits within the data pipeline.

Behavioural Descriptor To guide the QD search, we chose the Yet Another Mobile Network (YAMNet) DNN classifier to define our search space. The confidence scores output by the classifier for each class are used as selection signals for the QD algorithm, as discussed in section 3.1. While this pre-trained network

may limit our exploration, it was adopted in an effort to replicate a setup from previous evaluations of the Innovation Engine algorithm in the visual domain. That classifier is trained on AudioSet [9], which can be considered as a sonic sibling of the DNN classifiers trained on the ImageNet dataset [5]. YAMNet outputs 521 scores from a logistic (softmax) layer, corresponding to AudioSet classes. The classifier's output is intended "as a stand-alone audio event classifier that provides a reasonable baseline across a wide variety of audio events."[4]. Our approach to sound generation can be somewhat likened to a unique type of sound synthesiser, which is not crafted with the intention of mimicking natural sounds or creating textures that easily fit into well-known categories. Many modern generative models excel at such tasks [1], building on their prior training, but we considered the varied signal provided by this model as a good starting point for driving the EA towards diversity. We also considered it interesting to mirror the overall setup from experiments [28,26,18] that inspire our sonic investigations.

Periodic Signal Composition One factor potentially influencing the search space is our choice of CPPN activation functions and node types in the DSP graph. CPPNs have commonly been used to compose Gaussian, sigmoid, and periodic functions, such as in [35,34]. In our case, the pattern-producing network can only compose periodic functions, commonly used as oscillators in a variety of sound synthesis techniques: sine, square, triangle, and sawtooth. The node types in the DSP graph are the same as in [33], in addition to custom nodes, which were added to the repertoire in an effort to widen the search space. Those additional nodes are a wavetable and a specialised additive synthesis node, where multiple audio signals are sourced from the CPPN to fill a table in the former and represent partials or harmonics in the latter. The wavetable is traversed according to a control signal, also sourced from the CPPN, in a manner similar to vector synthesis. The partials in the additive synthesis node can be slightly inharmonic, according to a mutable parameter to each.

The duration of sounds rendered from each genome is defined by a linear ramp of values from -1 to 1 supplied to one CPPN input, while the pitch is controlled by the rate of a periodic (sine) signal at another input. Velocity is intended to simulate stimuli of different intensities when interacting with physical instruments, which is achieved by scaling the sine wave input by a velocity factor. The inputs are sampled at the same rate as the sampling rate of the audio graph.

QD Algorithm For the diversity-promoting algorithm, we chose Multi-dimensional Archive of Phenotypic Elites (MAP-Elites) [25]. Our experiments are based on a bespoke implementation of that algorithm, with the common addition of biasing it away from exploring niches that produce fewer innovations. This is achieved by assigning each niche a decrementing counter, representing a *curiosity score* as defined in [3] with constants set as in [18]. The counters start at a fixed value of 10, impacting the probability of that niche being selected for reproduction.

[4] YAMNet audio event classifier: https://tfhub.dev/google/yamnet

The classification outputs of the discriminative model define the cells of the behaviour space which the QD algorithm explores, where the performance at each niche is determined by the confidence values for each class. During our main runs of QD search, evaluations were performed in batches of 32.

Parameter Search Considering the temporal dynamics of sounds, and that the underlying pattern generator of our sound synthesis engine (CPPN) encodes sub-patterns that reveal over time, we performed preliminary experiments classifying sounds rendered at a different duration for each evaluation. One of the configurations involved 112 evaluations of each sound genome, rendering it to sounds of 4 durations, 7 pitch variations and 4 amplitudes. To explore other parameters of the QD search, such as mutation rates and their balance between the CPPN and DSP genome counterparts, as well as graph and node addition or deletion rates, we conducted a manual parameter search. Due to the computationally intensive nature of the task, these runs were based on a limited selection of parameter values. A comprehensive collection of plots from those runs can be found in the dataset accompanying this paper [15]. We found that evaluating sounds with a duration of half a second frequently led to the emergence of successful sound variants. Therefore, we decided to use this specific duration for assessing sounds in the QD runs of our primary experiments. Runs with node- and connection addition rates of 10% and corresponding deletion rates of 6% resulted in the best performance during our parameter search. As such, we ran with that as our baseline configuration along with equal probability of mutating each genome counterpart.

3 Results

For our main experiment, we ran 10 independent runs of the MAP-Elites algorithm, with the rates discussed in section 2 and behaviour evaluated by YAMNet on 0.5-second sounds. Each run lasted for 300 thousand iterations, with a batch or generation size of 32. At the start of each run, 50 seed iterations were performed, which differ from the rest of the iterations in that each individual is initialised from scratch rather than mutating a randomly selected elite occupying any of the cells.

3.1 Sound Generation- and QD Algorithm Variants

Aside from the set of evolution runs using our basic configuration described above, we performed two additional sets of runs. In one set, we altered the sound generation, and in the other set, we modified the progression of the QD algorithm.

Signal Processing Graph To investigate the impact of merging CPPNs with DSP graphs, we set up evolutionary runs in two distinct configurations: one in

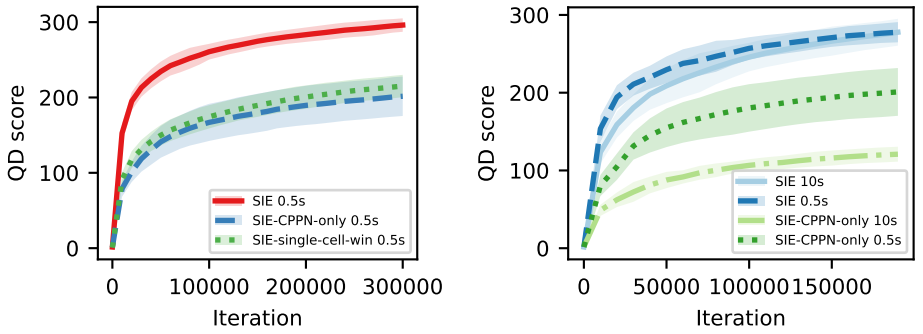


Fig. 2: On the *left* scores are plotted for the baseline configuration (section 2), evaluating sounds of 0.5s duration, along with variants where sounds are only rendered from CPPN mutations and where only one cell can be won at a time by each candidate elite. Data for each variant comes from 10 runs. The plot on the *right* compares the performance achieved when evaluating sounds rendered at two different durations—0.5s and 10s—from the baseline and CPPN-only run configurations, each independently executed 5 times.

which an evolved CPPN functioned solely as the audio signal source, providing a single output, and another where the CPPN was paired with an evolving DSP graph, allowing it to offer a multitude of audio and control signals, from up to 18 outputs.

In our experiments, we quantify the QD algorithm performance by calculating the QD-score [32,31]. This score is determined by summarising the confidence levels of the elites across the various classes delineated by YAMNet. When comparing the results from these runs, we observe in Figure 2 that the phenotypes (i.e., sound objects) produced from the genomes where CPPNs and DSP graphs were co-evolved achieved the highest overall QD-score. Through informal listening sessions conducted by the authors, it was observed that the sounds rendered from runs where the evolution of DSP graphs was allowed alongside CPPNs exhibited a higher degree of subjective aesthetic appeal. This phenomenon could potentially be attributed to the prevalence of classical synthesizer sounds, to which our ears have grown accustomed. In this context, the DSP graph can be seen as functioning akin to a modular synthesizer patch, rendering us less inclined to perceive the raw output generated by CPPNs as inherently pleasing. The rendered sounds can be auditioned in an online explorer (sec. 3.7) or accompanying dataset [15].

Behaviour Space Coverage The default behaviour of our MAP-Elites implementation allows each evaluated individual to win all cells where it performs better or where there is a vacancy, so it reaches full coverage from the first seed. To examine the effect of gradually covering the map of cells by allowing each

candidate to potentially win only one cell, the one where it receives the highest confidence from the classifier, we performed an identical set of runs except with that restriction in place. Runs where at most one cell at a time is won reached a coverage of 57.4%±3.4%, with their QD-score following a trajectory similar to that of full coverage CPPN-only runs, as depicted in figure 2, *left*.

Elite Populations Figure 3 (*left*) shows that the range of iterations where the current elites are found at the end of each run is sharply delimited around iterations 150K to 250K of the CPPN-only runs, while the CPPN+DSP runs continue to discover new elites more gradually throughout the latter half of the runs.

The set of unique elites at the end of CPPN-only runs is smaller than when co-evolving the DSP graphs, as plotted in figure 3 (*right*). Instead of distinguishing between individuals by their ID, where the differences could be only slight changes in e.g. connection weights, this plot is based on distinction between unique combinations of CPPN and DSP node and connection counts.

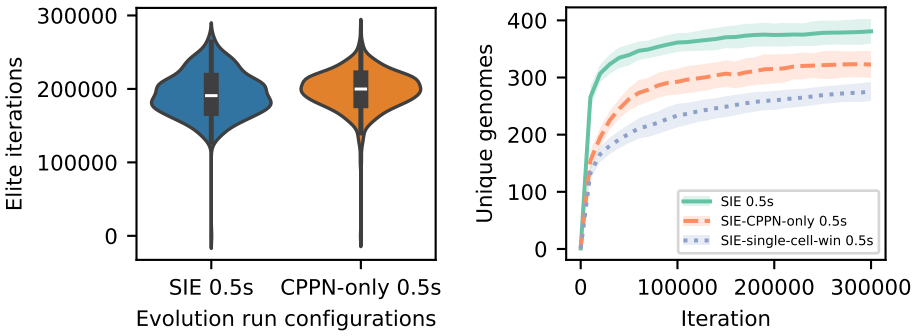


Fig. 3: *Left*: Distribution of iteration numbers at which the current class elite was discovered. *Right*: Count of unique individuals, as it evolves through iterations of the evolution runs.

3.2 Genome Complexity

The composition of audio graph nodes and CPPN activation functions can be seen in figure 4, where the prominence of the custom audiograph nodes (wavetable and additive synthesis, fig. 4, *bottom*) suggest that implementing other known techniques from the history of sound synthesis may be worthwhile. The distribution of CPPN activation function types is quite uniform in all variants of our runs (fig 4, *top*). It's also interesting to observe in the left plot of figure 6 that the CPPN-only runs resulted in more complex function compositions,

likely to compensate for the lack of a co-evolving DSP graph. This increased CPPN complexity resulted in longer rendering times and thus increased durations of the evolution runs, as that part of the genome is more computationally expensive, with potentially many network activations required for each sample, as discussed in [14].

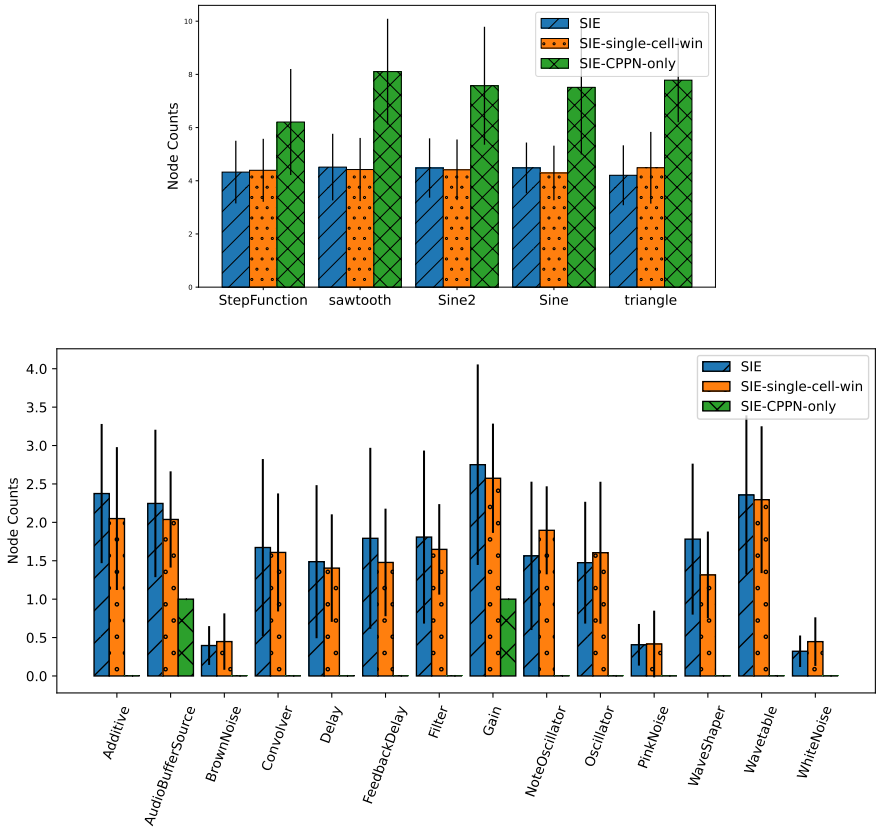


Fig. 4: Composition of CPPN activation functions (*top*) and DSP graph node types (*bottom*), from the different evolution run variants. It can be observed in the DSP chart that the CPPN-only variant does not evolve a DSP graph and only includes a source node for receiving the pattern-signal from the single CPPN output, and a gain node for passing it through to the output.

3.3 Performance Against Pre-trained Reward Signals

The YAMNet classifier chosen in this iteration of our investigations assigned high scores to the sounds generated by our system across most classes, as can

be seen in figure 5. There we can see again how the co-evolution of CPPNs with DSP graphs achieves higher scores overall. The figure also reveals how the synthesiser struggles in the range of classes between 214 and 276, which classify musical genres, rather than distinct sounds or instruments, such as “Pop music”, “Rhythm and blues”, “Flamenco”, etc. This is reasonable as the system is expected to generate sounds useful in the process of creating e.g. music, rather than entire musical compositions. Nonetheless it can be interesting to observe what the system came up with for those low-confidence classes, such as “Theme music”: a filter can be set in the online explorer (sec. 3.7) to audition classes containing the phrase “music” while scrubbing through the runs with a slider.

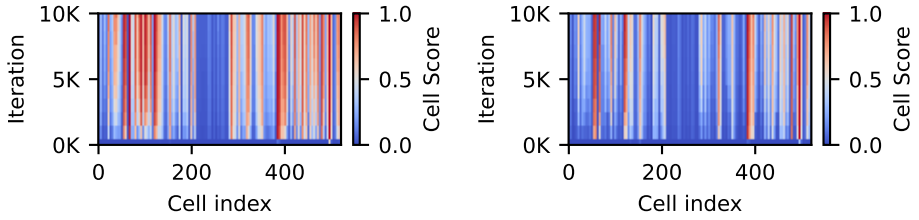


Fig. 5: Confidence scores declared by the YAMNet DNN, pre-trained on AudioSet classes (x-axis), averaged from the first 100 thousand iterations of 10 runs. Results from a set of runs where both CPPN and DSP genome counterparts are evolved can be seen on the *left* while the *right* map shows results from a set of runs restricted to evolution of the CPPN part of the genome, without evolving signal processing nodes.

3.4 Evolutionary Stepping-Stones

To assess how evolution leveraged the diversity promoted by our classifier, we conducted two measurements that explored the stepping stones across various classes. One has been called *goal switching* and defined as "the number of times during a run that a new class champion was the offspring of a champion of another class" in [28,26]. From our runs we measured a mean of 21.7±3.6 goal switches, 63.2% of the 34.3±4.5 mean new champions per class. This can be compared to the 17.9% goal switches reported in [26]. Another way of measuring how the evolutionary paths flow though the stepping stones laid out by the classifier is to trace through the phylogenetic tree leading to each elite and then count how often its parent comes from a class different from the one it occupies. Counting from the current elites of each class at the end of the evolution runs, we found a mean of 44.9±14.7 such occurrences. In lieu of a visual phylogenetic presentation, the *generation* slider of the evolution runs explorer (section 3.7) can dynamically reveal how elites for each class come from different, often unrelated classes during the course of evolution.

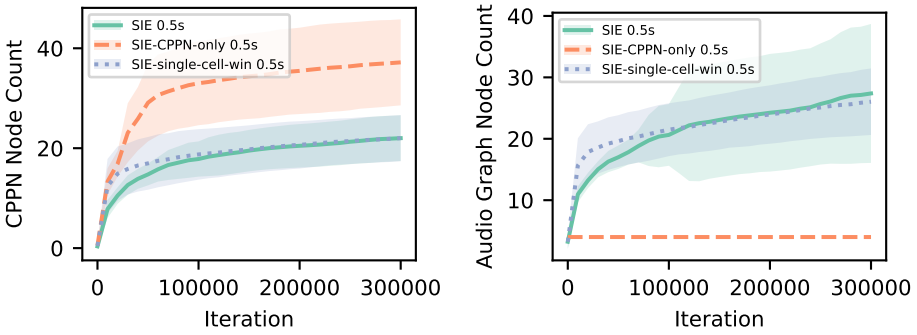


Fig. 6: Genome complexity over the course of 10 QD runs for each variant. CPPN node counts are plotted on the *left* and DSP graph node counts can be seen in the plot on the *right* (a flat line for the CPPN-only variant indicates that the DSP graph is not evolved.

3.5 Abandoning Diversity

Growth of genome complexity seems to have stayed within reasonable limits, even when CPPNs were left alone to the task of performing against the classifier (fig. 6). An exception to this is when we experimented with abandoning diversity and adopting single objectives. Though the benefit of diversity has been demonstrated [28,26], we investigated how a similar experiment fares in the sound domain. To that end, we selected 10 classes[5] as single objectives of separate runs and compared the performance and genome complexity with the performance from the QD runs on those same classes.

Interestingly, although the performance in single objective runs is higher on average than in multi-class runs, as shown in the *first* plot in figure 7, the difference is accompanied by a higher level of deviation and much higher genome complexity. The *second* and *third* plots in figure 7 indicate that the CPPN and DSP node counts in genomes from single objective runs are significantly higher than those of genomes from the same set of classes in QD runs. The computational effort required for the complex genomes evolved during the single class runs limited our iteration count to 50 thousand, 1/6th of the iterations performed for the baseline QD runs. The unexpected result of higher performance from the single-class runs may be attributed to the narrow set of chosen classes; this experiment could benefit from further investigation.

3.6 Temporal Pattern Revelation and Classifier Characteristics

Although half a second sounds were the most prevalent renditions of successful individuals in our manual parameters search (section 2), comparing sets of runs

[5] Single-class runs were performed on the classes Aircraft, Banjo, Beatboxing, Boom, Choir, Dubstep, Fusillade, Mandolin, Synthetic singing and Whistling.

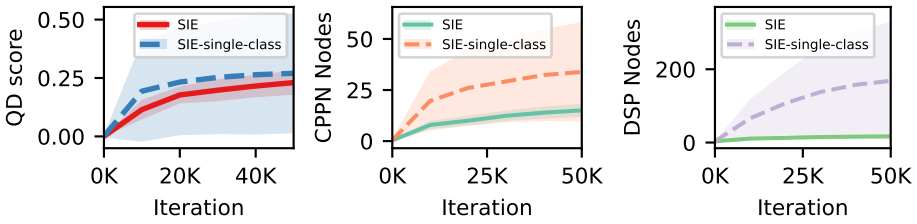


Fig. 7: The *first* plot shows performance scores from single-objective vs. multi-objective QD runs, averaged from a set of randomly selected classes. The *second* and *third* plots show how genome complexity developes during single- and multi-objective QD runs, in terms of CPPN and DSP graph node counts.

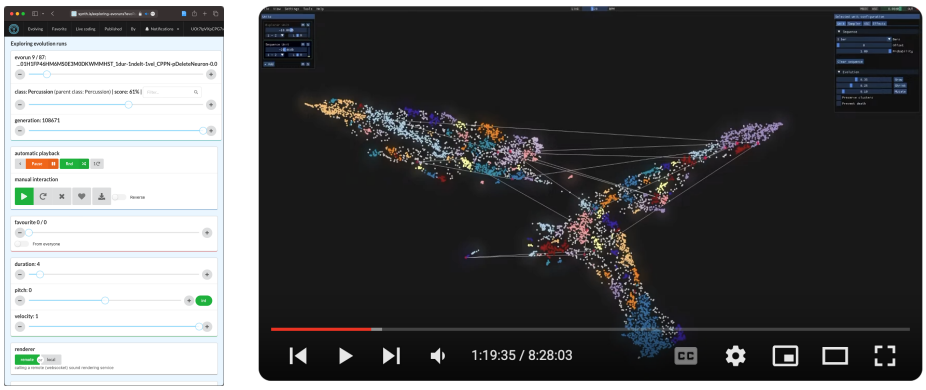
with two large variations in the duration of the evaluated phenotypes was interesting. We chose to compare runs evaluating half a second renditions of the evolved genomes with a set of runs evaluating ten-second renditions. One motivation for the choice of the longer duration, is that "YAMNet is trained on 1,574,587 10-second YouTube soundtrack excerpts from within ... AudioSet"[6]. While CPPN-only runs achieved less overall confidence when rendering 0.5s sounds for evaluation by the classifier, as can be seen on the *left* of figure 2, we hypothesised that allowing the classifier to sample in more detail the patterns developed by the CPPNs, when processing more frames over a longer duration, would result in higher confidence. The opposite turned out to be the case, where CPPN-only runs, rendering 10s sounds for evaluation achieved a lower QD score than corresponding runs rendering 0.5s sounds. Perhaps the lack of DSP becomes more significant in the evaluation of longer duration sounds. Duration has little effect when DSP graphs evolve alongside the CPPNs, as the *right* plot in figure 2 shows.

3.7 Access to Sound Objects and their Application

We have facilitated open access to the generated artifacts through different means. Those include an evolution runs explorer[7], depicted in figure 8a. Final elites from all runs have also been rendered to (128563) WAV files, which have been included in the accompanying dataset [15]. The sound objects in the pre-rendered files reflect the render-settings used to evaluate the corresponding genome that became an elite. The online explorer[7] provides greater flexibility as it dynamically renders sounds with the default settings, but the interface also enables users to modify these settings. This modification can potentially reveal other intriguing sonic behaviors from the same genome.

[6] YAMNet release announcement:
https://groups.google.com/g/audioset-users/c/U71MxTdHqkU
[7] Evolution runs explorer: https://synth.is/exploring-evoruns

As part of our investigation into the applicability of the discovered artefacts for creating other art, we loaded subsets of them into the experimental sampler AudioStellar [8] and used that software to drive evolutionary sequences through the phenotypes. A playlist of live-stream recordings showcasing evolutionary sequences using sounds discovered by QD runs is accessible online[8]. These compositions are largely automated, with human input limited to initial settings like evolutionary sequencing rates and fundamental sound effects. Nonetheless, they demonstrate the potential of the discovered sound objects to inspire creative endeavors. It is thought-provoking to consider if a human, given the same dataset, could craft more aesthetically pleasing arrangements with these sonic artefacts. We encourage the reader to obtain a copy of the files and engage in such experimentation [15].



(a) (b)

Fig. 8: (8a) Evolution runs explorer, where it is possible to scrub through evolution runs, their classes and generations. The sound properties duration, pitch and velocity can be changed and favourites can be collected. (8b) Live streams (recorded) of automated, evolutionary sequences through sounds rendered from the evolutionary runs discussed in this paper, as one way of experiencing and qualitatively evaluating the generated artefacts. The sequencing is peformed by the experimental sampler AudioStellar.

4 Conclusion and Future Work

Applying the combination of a diversity-promoting algorithm with selection pressure from a classifier reward signal to the search for sounds has been demonstrated to be a viable approach by the results discussed in this paper. Fur-

[8] Playlist with evolutionary sequences through sounds discovered by QD runs: https://youtube.com/playlist?list=PLSYAaR-xYhEXk0czfHYKJSWmZ8vG35xEN

thermore, the distinct approach to sound synthesis employed in this work has achieved high confidence from a DNN based classifier in most classes. High-scoring sounds are, in many cases, not the most realistic representatives of their class, especially when considering non-musical instrument classes, which can be attributed to how DNNs are easily fooled [27]. Other recently proposed classification approaches may be more robust and could be worth investigating [11,13], but classification robustness may not be the most sought-after quality in a creative system. With a focus on the diversity-promoting attribute of the selection pressure applied in this investigation, the diverse and innovative sound objects generated suggest that further explorations may be based on this system. The intent would be to broaden the range of potential discoveries within the sonic domain.

Adopting YAMNet as a classifier for sounds, to provide selection pressure for a QD algorithm, was a step towards investigating a simple version of the Innovation Engine algorithm in the domain of sounds. Further explorations may include expanding the behaviour space to search beyond predefined classes. This can be done by combining the feature extraction ability of a DNN, such as the one employed in this work, with dimensionality reduction (DR), as has been done in the visual domain in [22]. Extracting features with Variational Auto Encoders (VAEs) [17], and applying a clustering algorithm in the resulting latent space to define niches, as stepping stones during QD search, is another approach [23] worth exploring further in the domain of sounds. While VAEs require a training set, limiting the behaviour space to explore, periodically retraining a DR algorithm on discovered sound objects could enable autonomous and unsupervised discovery of the space of sounds which the generative system is able to render, without prior training, as proposed in [2,12]. Human intuition can also be leveraged to derive semantically meaningful diversity in the search space, as studied in [6], which can be especially important when generating sonic material leading to interesting discoveries according to individual aesthetics.

In the broadest sense, the concept of instruments has evolved from being a mere means to an end to a starting point for a journey into the unknown [21, p. 49]. The evolutionary system explored here is not intended as an instrument for serving requests from preconceived ideas but rather as a tool for discovering interesting sound objects that can steer the creative journey. The sound artefacts generated by our system, as discussed in this paper, are intended to facilitate or inspire the creation of further sonic art. This is different from the visual artefacts produced by many generative systems, which are often seen as standalone pieces without further utility. Instead of a top-down approach—where the end goals and characteristics of the desired sound are pre-defined—our method encourages a bottom-up process of exploration. This reflects the evolutionary path of human development, where cognitive skills have been shaped by the very tools that humans have uncovered. This echoes the saying, "the tool writes the toolmaker as much as the toolmaker writes the tool" ([4] as cited in [21, p. 5]). An instrument that promotes such exploratory discovery can enable us to continue on our path of evolution by developing human abilities through technology.

References

1. Choi, K., Im, J., Heller, L., McFee, B., Imoto, K., Okamoto, Y., Lagrange, M., Takamichi, S.: Foley Sound Synthesis at the DCASE 2023 Challenge. In: In arXiv e-prints: 2304.12521 (2023). https://doi.org/10.48550/arXiv.2304.12521
2. Cully, A.: Autonomous skill discovery with quality-diversity and unsupervised descriptors. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 81–89. ACM, Prague Czech Republic (Jul 2019). https://doi.org/10.1145/3321707.3321804
3. Cully, A., Demiris, Y.: Quality and Diversity Optimization: A Unifying Modular Framework. IEEE Transactions on Evolutionary Computation **22**(2), 245–259 (Apr 2018). https://doi.org/10.1109/TEVC.2017.2704781
4. Davis, W.: Replications : archaeology, art history, psychoanalysis. Pennsylvania State University Press (1996), iSBN: 0271015233 Place: University Park, Penn
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (Jun 2009). https://doi.org/10.1109/CVPR.2009.5206848, iSSN: 1063-6919
6. Ding, L., Zhang, J., Clune, J., Spector, L., Lehman, J.: Quality Diversity through Human Feedback (Oct 2023). https://doi.org/10.48550/arXiv.2310.12103, arXiv:2310.12103 [cs]
7. Gaier, A., Asteroth, A., Mouret, J.B.: Are quality diversity algorithms better at generating stepping stones than objective-based search? In: GECCO 2019 Companion - Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion. pp. 115–116 (2019). https://doi.org/10.1145/3319619.3321897
8. Garber, L., Ciccola, T., Amusategui, J.: AudioStellar, an open source corpus-based musical instrument for latent sound structure discovery and sonic experimentation. In: Proceedings of the International Computer Music Conference. pp. 62–67 (2021)
9. Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M., Ritter, M.: Audio Set: An ontology and human-labeled dataset for audio events. In: Proc. IEEE ICASSP 2017. New Orleans, LA (2017). https://doi.org/10.1109/ICASSP.2017.7952261
10. Godøy, R.I.: Chunking Sound for Musical Analysis. In: Ystad, S., Kronland-Martinet, R., Jensen, K. (eds.) Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music. pp. 67–80. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02518-1_4
11. Gong, Y., Lai, C.I.J., Chung, Y.A., Glass, J.: SSAST: Self-Supervised Audio Spectrogram Transformer (Feb 2022). https://doi.org/10.48550/arXiv.2110.09784, arXiv:2110.09784 [cs, eess]
12. Grillotti, L., Cully, A.: Unsupervised Behavior Discovery With Quality-Diversity Optimization. IEEE Transactions on Evolutionary Computation **26**(6), 1539–1552 (Dec 2022). https://doi.org/10.1109/TEVC.2022.3159855
13. Huang, P.Y., Xu, H., Li, J., Baevski, A., Auli, M., Galuba, W., Metze, F., Feichtenhofer, C.: Masked Autoencoders that Listen. In: NeurIPS (2022). https://doi.org/10.48550/arXiv.2207.06405
14. Jónsson, B.T., Erdem, C., Glette, K.: A System for Sonic Explorations with Evolutionary Algorithms. Journal of the Audio Engineering Society **72**(4) (2024). https://doi.org/10.17743/jaes.2022.0137
15. Jónsson, B.T., Glette, K., Erdem, C., Fasciani, S.: Supporting Data for: Towards Sound Innovation Engines Using Pattern-Producing Networks and Audio Graphs (2024). https://doi.org/10.18710/BAX9N5

16. Jónsson, B.T., Hoover, A.K., Risi, S.: Interactively Evolving Compositional Sound Synthesis Networks. In: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. pp. 321–328. GECCO '15, Association for Computing Machinery, New York, NY, USA (Jul 2015). https://doi.org/10.1145/2739480.2754796
17. Kingma, D.P., Welling, M.: Auto-Encoding Variational Bayes. In: Bengio, Y., LeCun, Y. (eds.) 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings (2014). https://doi.org/10.48550/arXiv.1312.6114
18. Lehman, J., Risi, S., Clune, J.: Creative Generation of 3D Objects with Deep Learning and Innovation Engines. In: Proceedings of the Seventh International Conference on Computational Creativity : ICCC 2016. pp. 180–187. 7, Sony CSL Paris, Paris, France (Jun 2016)
19. Lehman, J., Stanley, K.O.: Abandoning Objectives: Evolution Through the Search for Novelty Alone. Evolutionary Computation **19**(2), 189–223 (Jun 2011). https://doi.org/10.1162/EVCO_a_00025, conference Name: Evolutionary Computation
20. Lehman, J., Stanley, K.O.: Evolving a diversity of creatures through novelty search and local competition. Genetic and Evolutionary Computation Conference, GECCO'11 (Gecco), 211–218 (2011). https://doi.org/10.1145/2001576.2001606, iSBN: 9781450305570
21. Magnusson, T.: Sonic writing: technologies of material, symbolic and signal inscriptions. Bloomsbury Academic, New York, NY (2019)
22. McCormack, J., Cruz Gambardella, C.: Quality-Diversity for Aesthetic Evolution. In: Martins, T., Rodríguez-Fernández, N., Rebelo, S.M. (eds.) Artificial Intelligence in Music, Sound, Art and Design. pp. 369–384. Lecture Notes in Computer Science, Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-03789-4_24
23. McCormack, J., Gambardella, C.C., Krol, S.J.: Creative Discovery using QD Search (May 2023). https://doi.org/10.48550/arXiv.2305.04462, arXiv:2305.04462 [cs]
24. Moore, B.C.: Hearing. Academic Press (1995), iSBN: 0125056265 Place: San Diego, Calif Series: Handbook of perception and cognition (2nd ed.)
25. Mouret, J.B., Clune, J.: Illuminating search spaces by mapping elites (Apr 2015). https://doi.org/10.48550/arXiv.1504.04909, arXiv:1504.04909 [cs, q-bio]
26. Nguyen, A., Yosinski, J., Clune, J.: Understanding innovation engines: Automated creativity and improved stochastic optimization via deep learning. Evolutionary Computation **24**(3), 545–572 (Sep 2016). https://doi.org/10.1162/EVCO_a_00189
27. Nguyen, A., Yosinski, J., Clune, J.: Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. arXiv (Apr 2015). https://doi.org/10.48550/arXiv.1412.1897, arXiv:1412.1897 [cs]
28. Nguyen, A.M., Yosinski, J., Clune, J.: Innovation Engines: Automated Creativity and Improved Stochastic Optimization via Deep Learning. In: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. pp. 959–966. GECCO '15, Association for Computing Machinery, New York, NY, USA (Jul 2015). https://doi.org/10.1145/2739480.2754703
29. Nordmoen, J., Veenstra, F., Ellefsen, K.O., Glette, K.: MAP-Elites enables powerful stepping stones and diversity for modular robotics. Frontiers in Robotics and AI **8**, 56 (2021). https://doi.org/10.3389/frobt.2021.639173

30. Noë, A.: The entanglement : how art and philosophy make us what we are. Princeton University Press, Princeton, New Jersey (2023), iSBN: 9780691188812 Place: Princeton, New Jersey
31. Pugh, J.K., Soros, L.B., Szerlip, P.A., Stanley, K.O.: Confronting the Challenge of Quality Diversity. In: Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation. pp. 967–974. GECCO '15, Association for Computing Machinery, New York, NY, USA (Jul 2015). https://doi.org/10.1145/2739480.2754664
32. Pugh, J.K., Soros, L.B., Stanley, K.O.: Quality Diversity: A New Frontier for Evolutionary Computation. Frontiers in Robotics and AI **3** (Jul 2016). https://doi.org/10.3389/frobt.2016.00040
33. Rice, D.: GenSynth: Collaboratively Evolving Novel Synthetic Musical Instruments. Master's thesis, The University of Oklahoma (May 2015). https://doi.org/10.13140/RG.2.1.4691.6001
34. Secretan, J., Beato, N., D'Ambrosio, D.B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J.T., Stanley, K.O.: Picbreeder: a case study in collaborative evolutionary exploration of design space. Evolutionary Computation **19**(3), 373–403 (2011). https://doi.org/10.1162/EVCO_a_00030
35. Stanley, K.O.: Compositional pattern producing networks: A novel abstraction of development. Genetic Programming and Evolvable Machines **8**(2), 131–162 (Jun 2007). https://doi.org/10.1007/s10710-007-9028-8
36. Stanley, K.O., Miikkulainen, R.: Evolving Neural Networks through Augmenting Topologies. Evolutionary Computation **10**(2), 99–127 (Jun 2002). https://doi.org/10.1162/106365602320169811
37. Takagi, H.: Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation. Proceedings of the IEEE **89**(9), 1275–1296 (Sep 2001). https://doi.org/10.1109/5.949485, conference Name: Proceedings of the IEEE
38. Wyse, L.: Free music and the discipline of sound. Organised Sound **8**(3), 237–247 (Dec 2003). https://doi.org/10.1017/S1355771803000219