CONVERSING FOR COEXISTENCE

Patrik LECHNER, **Marco DÖTTLINGER**, and **Thomas GRILL**

mdw – University of Music and Performing Arts Vienna, Wien, Austria

ABSTRACT

The proposed audio installation work explores emergent relations between distinct agents – human, technological, and environmental – within an adaptive, electroacoustic ecology. Sound becomes a medium for attuning to difference, negotiating tension, and practicing or simulating coexistence, thereby enabling an agent's potential of finding their individual niches within the system. To achieve this, we developed a tool set encapsulating expression, agency and perception of our machine agents. While the technical setup is presented in a prototypical form, we are contributing a small but versatile tool that allows for integrating reinforcement learning algorithms and multi-agent setups in common sound or media art workflows via Open Sound Control. The project foregrounds adaptive algorithms and dynamical adaptive systems in both machine listening and synthesis as aesthetic strategy and proposes a performative ecology and ethical practice. We claim that the installation does not merely generate sounds; it enacts a *speculative sociality* – a sonic experiment in how heterogeneous beings might inhabit shared worlds without erasure but by 'Conversing for Coexistence', by *making kin*.

1. CONTEXTUALIZATION

This installation is embedded in a multi-year project *Spirits in Complexity – Making kin with Experimental Music Systems* (launched April 2024, a cooperation between the mdw Vienna, JKU Linz, KTH Stockholm, and the Intelligent Instruments Lab). In this project, affective relationships between human actors and musical systems are at the center of our attention. We see music as an emergent process, rising from the interaction of humans and sounding instruments within social, cultural and architectural contexts, often fostering close functional and affective relationships with acoustic instruments, recording equipment, synthesizers and software.

1.1 Preliminary Work

Our preceding *Rotting Sounds* [1] project conducted extensive artistic experiments with feedback systems [2–4], where individual sounding entities interweaved into an integral soundscape within the *Auditorium of Rotting Sounds* [5]. This exemplifies what Di Scipio [6, p.49] terms *structural coupling*: "computing becomes neither an entirely deterministic process, nor an indeterministic one, but a driving active part of a larger complex system." In such recurrent audio installations, even tiny mutual interactions are exponentially amplified over iterations [1], forming complex networks of cross-resonant entities – akin to Di Scipio's audible ecosystems [7] and Tudor's *Rainforest* series [8]. The proposed work extends this into the realm of *non-trivial* instruments [9, 10], where instrument, space, audiences, and algorithms co-create music as a continuously emergent phenomenon. Beyond our own prior work, the installation draws on the acoustic niche hypothesis [11], which describes how species partition a shared soundscape by frequency and time to avoid masking one another, and its extension into generative art as *niche construction* [12], where agents actively reshape their environment rather than merely occupying a fixed slot within it – a closer fit to our agents' behavior than niche-occupation alone. On the technical side, our multi-agent setup connects to work on emergent communication in reinforcement learning, where agents develop coordinated signaling strategies without being explicitly programmed to do so [13].

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. CONVERSING FOR COEXISTENCE

The presented audio installation work "Conversing for Coexistence" explores the emergent relations between distinct agents – human, technological, and environmental – within an adaptive, electroacoustic ecology.

The project foregrounds adaptive algorithms in both machine listening and synthesis as aesthetic strategy and proposes a performative ecology and ethical practice: autonomous yet interdependent sonic entities that continuously negotiate their presence within a shared acoustic environment.

As such, multiple realizations in hardware are possible, and as this is an ongoing project we are actively exploring these possibilities. One such version was recently presented at the Conference of the Society for Artistic Research (SAR) in Galway [1]. There, a setup of two agents was used, 3D-printed boxes equipped with a speaker, microphones and a Raspberry Pi single board computer. In this system, each agent is listening to its surrounding, the audience, and the environment as a whole and contributes a sound, actively trying to find a spectral and temporal niche in an ongoing negotiation. Audience noises are perceived

[1] `https://sar2026.ie/conversing-coexistence`, accessed 2026-07-10

the same way as other agents or the environment – no artificial hierarchy is introduced. During the Galway presentation, this prompted audience members to imitate or provoke specific agents, attempting to claim or negotiate a niche for themselves.

Niche-finding, discussed as *acoustic niche hypothesis* [11, 14, 15] becomes the main organisation model and is both a technical procedure and an aesthetic principle.

In this sense, sound becomes a medium for attuning to difference, negotiating tension, and practicing or simulating coexistence, with a strong emphasis on *making kin*: novel sonic artifacts arise through permanent friction and relational interplay rather than control. These *resonant* relationships between human and non-human actors can take the form of kinship, negotiation, or ritual. Following [16], *making kin* denotes non-biological, non-hierarchical relations enacted through shared practices, mutual responsiveness, and co-habitation of worlds – making acoustic communication central. The sounding agents invite the audience to communicate, to 'play with them' more than 'playing them', as a more traditional instrument's affordances might suggest. A different way of framing this is to again connect this to Haraway's ideas: sympoiesis instead of autopoiesis. Sympoiesis as *making-with* could be seen as the central theme of this installation as the result is mainly interesting in its interplay of human and non-human agents. As a consequence, staying in Haraway's framework, the natural mode of engaging with the project as technological or human agent is described as making-kin. All participating agents enact a *speculative sociality* – an experiment in how heterogeneous beings might inhabit shared sound worlds without erasure, masking or overriding, but by finding their individual niches.

In ecological terms, a niche is not merely a place but a relational position – defined by difference, interdependence, and constraints. Transposed into our electroacoustic or sonic arts practice, this means that each agent learns to articulate itself through contrast and accommodation. In other words, when the sonic field becomes too saturated, agents must retreat, transform, or even risk extinction: They have to radically change their sonic behavior to regain their niche.

Novelty arises not from control or simple randomness alone, but from this ongoing negotiation. The system never stabilizes into a fixed composition; it remains in a state of dynamic disequilibrium. Sonic events are co-produced through micro-adjustments, misalignments, and emergent alliances between agents. The installation thus performs an ecology in real time: a field of distributed listening and response.

2.1 Technical Aspects

The following shall give a brief overview of our implementation as well as a technical contextualization. A prototypical implementation in the Max/MSP and Python programming environments can be found on GitHub [17, 18]. All parts of the installation – synthesis, feature extraction, and reward function – follow an admittedly simplistic approach. We believe this simplicity is a strength, as it allows emergent sympoietic phenomena to remain in the foreground; however, the current implementation can only gesture toward the described phenomena, such as negotiation or social activity, without any claim to modeling them accurately or completely. To implement a simple form of negotiation and niche-finding, each agent is rewarded for maximizing the distance between its own feature vector and that of its environment.

2.1.1 State of the Art

At the core of our installation lies a reinforcement learning (RL) based control mechanism, implemented in the Python programming language. While RL has been used in many music specific applications, few works seem to make use of these techniques quite in the way implemented here. Multiple works in the past have employed RL on a symbolic level such as MIDI for example [19], [20] or [21]. RL from Human Feedback (RLHF) has for example been used in [22], but here it was also used in order to strengthen text-adherence of the musical output.

Broadly speaking, RL was often applied in order to create music generation machines aligned with human aesthetic goals. On the contrary, our approach sees the design of a reward function in combination with feature extraction as an act of composition. In that aspect, our approach is more comparable with projects such as the 'wekinator' [23] or 'flucoma' [24] , where dataset creation, learning and mapping are all thought of as being more integrated into an artistic or experimental practice. Finally, simplified social interaction models have a long history, be it based on physically inspired models [25] or rule based systems [26] and our work connects modern machine learning techniques to this tradition of exploring emergent patterns artistically.

2.1.2 Agent Architecture

The RL community has built various tools in order to train agents, such as 'gymnasium' [27]. Building on top of this given structure we have developed a tool that allows easy integration of custom reward functions and agent training via Open Sound Control (OSC). Training is currently performed via the 'stable-baselines' Python package [28]. The tool we developed is available online on GitHub [18]. As depicted in Figure 2, the Python based tool receives observations and rewards over OSC and sends out actions over OSC. To achieve our goal of creating a niche-finding behavior, our observation vector consists of both an analysis of the environment and the agent's own output. Feature extraction is performed by what is called $F(x)$ in Figure 2 and described below in Section 2.1.3. As one of our goals was to implement the whole structure on very light hardware and observe the emergent behaviors of our collective of agents rather than to control the situation, many technical choices were made with simplicity in mind. The reward function, $R(\vec{o})$ in Figure 2, is the Euclidean distance between the two feature vectors. Some scaling and a slight penalty for extreme parameter settings/actions was introduced in order to enforce more interesting dynamics. As noted above, using the Euclidean distance between low-level features is a minimal proxy for niche differentiation.

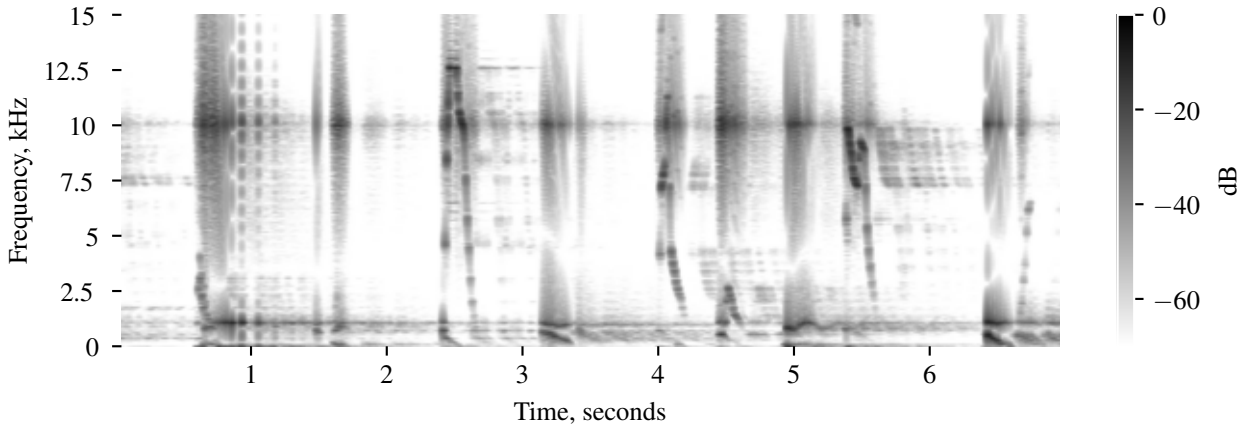


Figure 1. Example spectrogram of two agents interacting over seven seconds.

The model being optimized during the run of our installation is `stable-baselines3`'s `A2C` model based on [29]. As the agents learn in real time, multiple parameters of the learning process can be used to influence the aesthetic results, such as the learning rate or the agent's speed (i.e., the step size in changing synthesis parameters). Even with this simplistic example setup and only two agents rather complex spectral movements and glissandi can be observed, as can be seen in Figure 1.

For testing and demonstration reasons, we have implemented an even simpler version of the setup, in which the synthesis engine is reduced to a source of white noise filtered with a bandpass filter. In this very simplified setup, the agent can only control the frequency of the filter. Figure 4a shows these two simplified agents freely interacting. One of the two agents has settled for a very low frequency, barely visible in the bottom spectrogram, while the other one frequently restarts new learning episodes but has a tendency towards high frequencies. With further training, this system settles for a steady-state where maximum separation is achieved, and no further optimization is observed.

As a further test, we disturbed this simplified setup with environmental noise. Both agents receive white noise filtered with a 3rd order highpass filter at 8 kHz. Expectedly, the system takes more time to stabilize and tends towards lower frequencies, still keeping separation in a smaller range, see Figure 4b.

2.1.3 Feature Extraction

Due to hardware limitations and the prototypical nature of the installation, feature extraction is done purely in the time domain. In order to give the agents the possibility to find their environmental niche, only three simple features are used: The *RMS* energy, approximated via an IIR filter, the zero crossing rate and the spectral centroid, which can for example be calculated in the time domain as shown in [30], and as it is implemented in the FAUST libraries. In our current prototype, we are using the `zsa.Descriptors` [31] implementation of the spectral centroid.

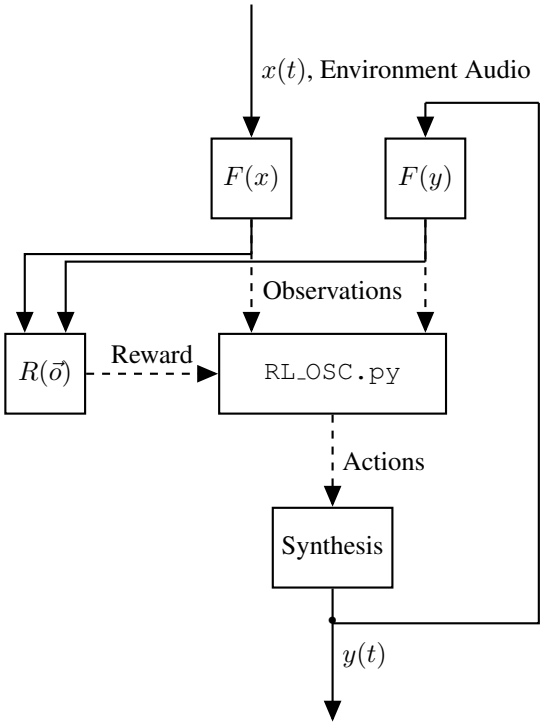


Figure 2. Diagram of one agent. Solid lines are audio rate, dashed lines are UDP/OSC connections. $F(x)$ performs audio feature calculation, $R(\vec{o})$ computes a reward from a pair of observation vectors.

2.1.4 Synthesis

As of now, two prototypical synthesis engines have been explored. The aforementioned simplistic single resonator as a demonstration on the one hand and a richer but still preliminary experiment on the other. For this latter more interesting synthesis engine we explored a technique that is aesthetically inspired by David Tudor's *Rainforest* installation and uses a form of bird song synthesis derived from [32]: A square wave oscillator with very narrow pulse width feeding three parallel resonators, see Figure 3. The nature of this current synthesis should be seen as a rem-

iniscence; the final realization of the project will introduce sonically richer material and a more complex feature extraction in order to facilitate more gestural expression of the agents. Still, given this extension and a total of eight parameters for each agent to act upon, already a fair amount of sonic and rhythmic complexity can be observed.

Audio examples of the current state can be listened to on a respective Research Catalogue exposition [2] .

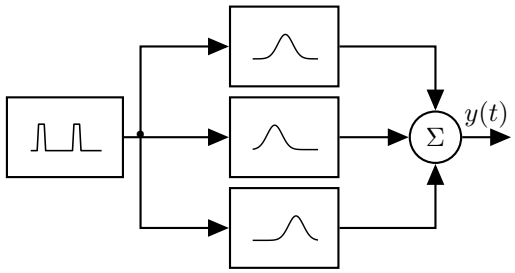


Figure 3. Preliminary synthesis engine.

2.2 Collective Agency

At its core our installation piece is a distributed system of synthesis devices – speaker-agents equipped with reinforcement learning capacities [18]. Each agent dynamically listens and adapts to the sonic environment it cocreates. Rather than optimizing toward a fixed compositional goal, it adapts its behavior in response to all other participating sound-making entities, the space it is living in and other (human) interaction. Each agent does not seek for domination, replication, or harmonization, but to continuously adapt, discover and maintain a viable niche: a permanently shifting auditory space where it can exist without erasing the presence of other sounding beings. Each speaker-agent operates with partial knowledge and limited perspective. It does not access a total view of its world; instead, it perceives locally and acts accordingly. Collective form on any level emerges from these situated decisions.

What arises is a densely interwoven sonic work of mutually influencing sound sources – a network of entities that communicate through and with one another. In this sense, this work can be understood as an ecosystem of audible relations, where sound circulates across bodies and objects rather than being linearly produced and consumed. It recalls earlier practices in which sounding objects, circuits, and spaces were coupled into responsive constellations, and where the acoustic outcome was the result of systemic interaction rather than individual authorship (Tudor's Rainforest series and Di Scipio's audible ecosystems). What Di Scipio would call structural coupling, we are approaching as sympoiesis and kinship, offering an emphasis on communication and making-with.

The installation space is not a neutral container but an active participant of an electroacoustic ecology. Architectural resonances, audience movement and their sound production, and ambient noise can also feed back into the adaptive processes. The acoustic environment functions as both habitat and co-agent. Through this feedback loop, the work stages an electroacoustic ecology in which boundaries between instrument, audience, and environment become porous. Sound is not projected into space; it circulates within it. The agents learn from the reverberant traces of their own actions and from the responses of others, producing a continuously evolving soundscape that reflects the specific conditions of its site. Each iteration of the work is therefore singular. Such an ecology cannot be fully replicated, because it depends on the contingencies of place and presence.

As argued above, *Making kin* in this context is to cultivate durable yet mutable relations across difference. The adaptive agents do not merge into homogeneity; they maintain distinct sonic identities while remaining responsive to the collective, the environment. Kinship here is based on sustained negotiation and entails a shift in listening practice. Our installation does not simply illustrate kinship, it enacts its dynamic nature: "We are entangled; we are processes of becoming; we are products of resistance to the conditions in which we find ourselves" [33, p. 139].

This perspective radically transforms the role of the participants. Instead of seeking narrative progression or authorial intention, listeners are invited to attune to patterns of emergence. Attention becomes a mode of cohabitation. When alliances form and dissolve or tensions accumulate, we are participating in the constant self- and reorganization of a shared organism. In this state of entanglement, the work's self-willed dynamics – its resistances, instabilities, and occasional breakdowns – are not seen as errors but as expressions of agency, points of friction where the boundary between the observer and the environment blurs. Engaging with these behaviors requires patience and reciprocity rather than correction. In this sense, making kin is both an artistic strategy and an ethical orientation toward the complexity of our inseparable existence with the world.

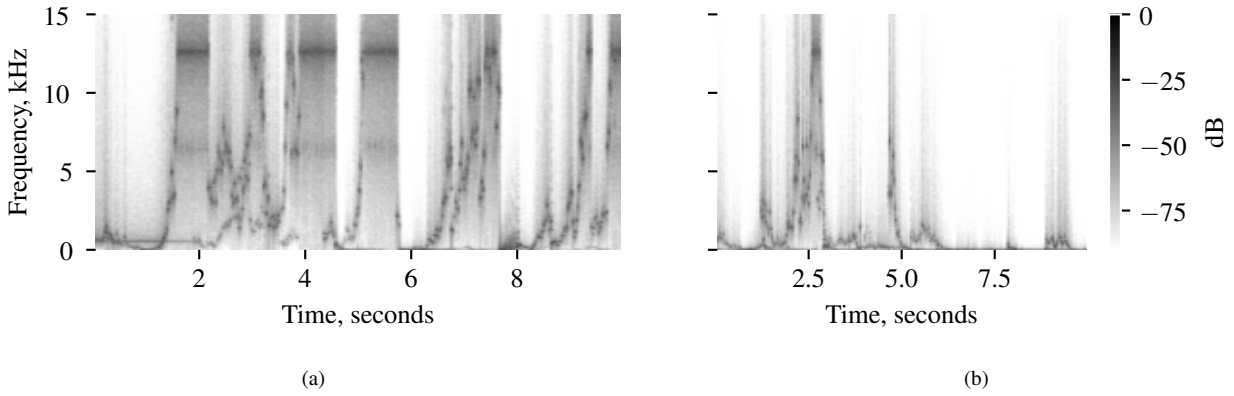
Niche-finding as the main compositional logic contrasts dominant paradigms of optimization and control in technological culture. It re-imagines artistic intelligence as ecological: a capacity to sustain multiplicity, to negotiate space, and to produce newness through relational becoming. We claim that the installation does not merely generate sounds; it enacts a *speculative sociality* – a sonic experiment in how heterogeneous beings might inhabit shared worlds without erasure but by *Conversing for Coexistence*.

3. CONCLUSION

We present an installation exploring themes of social negotiation for coexistence, exploring the concept of the acoustic niche. We show that a reinforcement learning based implementation allows for clear definition of basic agentic and social behaviors and contribute a small versatile workflow that allows the wider community easy access to such strategies regardless of their preferred software environments.

Given the background of our research project that centers around making kin with experimental music machines, we presented both the theoretical background and the ad-

[2] https://www.researchcatalogue.net/view/4309614, accessed 2026-03-27



(a) (b)

Figure 4. Comparison of spectrograms of two runs with a simplified synthesis engine and two agents interacting. (a) shows the agents interacting freely over the whole frequency range. (b) is created with environmental noise present in the high frequency range and displays tendencies towards low frequency content.

vancements of this contribution specifically as it employs thinking about agencies as a form of composition and represents a different way of designing an evolving experimental music setup in connection with human actors and a sonic environment.

Acknowledgments

This research was funded in whole or in part by the Austrian Science Fund [Grant DOI: 10.55776/AR821]. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

4. REFERENCES

[1] T. Grill, T. Bovermann, and A. Schilling, "Rotting sounds – artistic research practice in experimental sound art," *Zeitschrift der Gesellschaft für Musiktheorie*, vol. 19, no. 2, pp. 103–118, 2022.

[2] K. Filip and T. Grill, "Dust a bit," Audio installation, 2019. [Online]. Available: https://rottingsounds.org/threads/auditorium/dust-a-bit/

[3] M. de Vega, "Intermission," Audio installation, 2019. [Online]. Available: https://rottingsounds.org/threads/auditorium/intermission/

[4] D. Sanfilippo, "Phase transitions," Audio installation, 2019. [Online]. Available: https://rottingsounds.org/threads/auditorium/phase-transitions/

[5] T. Grill, T. Bovermann, A. Schilling, and T. Leibetseder, "Auditorium of rotting sounds," Exhibition catalogue, Vienna, Austria, March 2019. [Online]. Available: http://doi.org/10.5281/zenodo.19267046

[6] A. Di Scipio, "The place and meaning of computing in a sound relationship of man, machines, and environment," in *Proceedings ICMC—SMC*, Athens, Greece, September 2014.

[7] R. Meric and M. Solomos, "Analysing audible ecosystems and emergent sound structures in di scipio's music," *Contemporary Music Review*, vol. 33, no. 1, pp. 4–17, 2014.

[8] J. Driscoll and M. Rogalsky, "David tudor's rainforest: an evolving exploration of resonance," *Leonardo Music Journal*, vol. 14, pp. 25–30, 2004.

[9] T. Grill, A. Castelló, P. Lechner, M. Döttlinger, and A. Flexer, "Black box music – performing non-trivial instruments," *ECHO*, vol. 7, December 2025. [Online]. Available: https://echo.orpheusinstituut.be/article/black-box-music

[10] M. Döttlinger and T. Grill, "Spirits in Complexity – Intra-Aktionen mit Algorithmen," *Neue Zeitschrift für Musik*, no. 1, 2025.

[11] B. L. Krause *et al.*, "The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats," *The Soundscape Newsletter*, vol. 6, no. 5, 1993.

[12] J. McCormack and O. Bown, "Life's what you make: Niche construction and evolutionary art," in *Workshops on Applications of Evolutionary Computation*. Springer, 2009, pp. 528–537.

[13] I. Mordatch and P. Abbeel, "Emergence of grounded compositional language in multi-agent populations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[14] A. Anjana and K. Teji, "Ecology of acoustic partitioning in insect assemblages," *Behavioural Processes*, vol. 226, p. 105167, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0376635725000294

[15] V. García-Navas, N. Feliu, and D. T. Blumstein, "Changes in the acoustic structure of australian bird communities along a habitat complexity gradient," *Behavioral Ecology*, vol. 34, no. 6, pp. 930–940, 2023.

[16] D. J. Haraway, *Staying with the Trouble*. Duke University Press, 2016.

[17] P. Lechner, "hrtlacek/conversing_for_coexistence_maxpatches: 1.0," Mar. 2026. [Online]. Available: https://doi.org/10.5281/zenodo.19236817

[18] ——, "hrtlacek/rl-osc: Alpha," Mar. 2026. [Online]. Available: https://doi.org/10.5281/zenodo.18827080

[19] N. Jonason, L. Casini, and B. L. Sturm, "Smart: Tuning a symbolic music generation system with an audio domain aesthetic reward," *arXiv preprint arXiv:2504.16839*, 2025.

[20] N. Jaques, S. Gu, R. E. Turner, and D. Eck, "Tuning recurrent neural networks with reinforcement learning," 2017.

[21] N. Kotecha, "Bach2bach: Generating music using a deep reinforcement learning approach," *arXiv preprint arXiv:1812.01060*, 2018.

[22] G. Cideron, S. Girgin, M. Verzetti, D. Vincent, M. Kastelic, Z. Borsos, B. Mcwilliams, V. Ungureanu, O. Bachem, O. Pietquin, M. Geist, L. Hussenot, N. Zeghidour, and A. Agostinelli, "MusicRL: Aligning music generation to human preferences," in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp, Eds., vol. 235. PMLR, 21–27 Jul 2024, pp. 8968–8984. [Online]. Available: https://proceedings.mlr.press/v235/cideron24a.html

[23] R. Fiebrink and P. R. Cook, "The wekinator: a system for real-time, interactive machine learning in music," in *Proceedings of The Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010)(Utrecht)*, vol. 3. Citeseer, 2010, pp. 2–1.

[24] P. Tremblay, O. Green, G. Roma, J. Bradbury, T. Moore, J. Hart, and A. Harker, "The fluid corpus manipulation toolbox (v. 1). zenodo," 2022.

[25] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 1987, pp. 25–34.

[26] J. Conway, "Conway's game of life," *Scientific American*, vol. 223, pp. 120–123, 1970.

[27] M. Towers, A. Kwiatkowski, J. Terry, J. U. Balis, G. De Cola, T. Deleu, M. Goulão, A. Kallinteris, M. Krimmel, A. KG *et al.*, "Gymnasium: A standard interface for reinforcement learning environments," *arXiv preprint arXiv:2407.17032*, 2024.

[28] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu, "Stable baselines," https://github.com/hill-a/stable-baselines, 2018.

[29] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. P. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," 2016. [Online]. Available: https://arxiv.org/abs/1602.01783

[30] D. Sanfilippo, "Time-domain adaptive algorithms for low-and high-level audio information processing," *Computer Music Journal*, vol. 45, no. 1, pp. 24–38, 2021.

[31] M. Malt and E. Jourdan, "Zsa. descriptors: a library for real-time descriptors analysis," in *5th Sound and Music Computing Conference, Berlin, Germany*, 2008, pp. 134–137.

[32] B. Schottstaedt, *Common Lisp Music Documentation*, Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, 1992, accessed February 20, 2026. [Online]. Available: https://ccrma.stanford.edu/software/snd/snd/clm.html

[33] A. Noë, *The Entanglement: How Art and Philosophy Make Us What We Are*. Princeton: Princeton University Press, 2023.