Sound Swarm: A Synthetic Ecology of Embodied Mesh Synthesizers for Emergent Soundscapes

Mikhail Mansion
The University of Tokyo
Tokyo, Japan
mikhail@xlab.iii.u-tokyo.ac.jp

Yasuaki Kakehi
The University of Tokyo
Tokyo, Japan
kakehi@iii.u-tokyo.ac.jp

Figure 1: Concept rendering (not a photograph) of *Sound Swarm* in an illustrative environment.

Abstract

Sound Swarm is an embodied, multi-agent system that functions as a self-organizing mesh synthesizer, generating soundscapes through distributed sensing and environmentally mediated interaction. As these systems scale, traditional musical concepts—individual voice assignment, score-based synchronization, and deterministic event mapping—break down because the acoustic environment introduces propagation delay, masking, reflections, and spatial heterogeneity. We argue that composing for such systems requires a shift from event-specification to *condition-setting*: designing interaction rules, constraints, spatial arrangements, and time-scales under which sonic organization can emerge. This paper details a three-tier architecture—Perception, Behavior, and Expression—that decouples real-time audio synthesis from low-rate "colony cognition." We introduce a graph-based method for collective spatial sense-making: pairwise RF/acoustic measurements populate a relational tensor which is interpreted by a lightweight neural relational model to infer topology in situ. The framework is grounded in a biological taxonomy of coordination behaviors: honeybee-inspired role election for hierarchy, ant-inspired stigmergy for environmental coupling, and tree frog models for rhythmic coordination. By treating the swarm as a synthetic ecology, we demonstrate how emergent musical form arises from the triadic interaction between user-defined conditions, the physical environment, and agent-level biological heuristics.

NIME '26, June 23–26, 2026, London, UK

CCS Concepts

• **Applied computing** → **Sound and music computing**; • **Computing methodologies** → **Multi-agent systems**; • **Human-centered computing** → *Interactive systems and tools.*

Keywords

Embodied Interaction, Generative Soundscapes, Swarm Intelligence, Acoustic Ecology, Decentralized Systems, Topology Inference

ACM Reference Format:
Mikhail Mansion and Yasuaki Kakehi. 2026. Sound Swarm: A Synthetic Ecology of Embodied Mesh Synthesizers for Emergent Soundscapes. In *Proceedings of International Conference on New Interfaces for Musical Expression (NIME '26).* ACM, New York, NY, USA, 13 pages.

1 Introduction

Emergent phenomena—where coherent global structure arises from simple local interaction rules—have long served as a conceptual bridge between biology, computation, and music. From the cellular automaton known as the *Game of Life* [10] to Reynolds' flocking model [21], such systems demonstrate how complex organization can arise without centralized control. In computer music, these ideas informed swarm-based approaches, most notably in the work of Blackwell and colleagues, who showed that musical structure can emerge through decentralized agent interaction rather than explicit compositional specification [3, 4].

In biological systems, self-organization arises not from global representations but from local sensing, indirect environmental interaction, and adaptive response thresholds. The concept of *stigmergy*, originally introduced by Grassé [12], describes how modifications to a shared environment can guide collective behavior without explicit communication. Work on cognitive stigmergy further emphasizes that collective structure can emerge under partial and noisy information, where the medium itself functions as shared memory and a coordination substrate rather than a neutral channel [22].

These principles have strong parallels in music and sound practice. Ecosystemic approaches frame sound as both signal and environment, emphasizing feedback between sonic action and physical context [7, 23]. Behavioral models of musical agents describe sound-producing entities whose output emerges from interaction rather than top-down control [5]. Together, these perspectives position musical interaction as situated, embodied sense-making rather than symbolic execution.

Here, we use embodiment to refer to systems in which sound-producing agents are physically co-present and interact through real acoustic propagation, such that the environment participates in shaping interaction. While prior swarm-music systems often operate in simulated or abstracted spaces, this work explores a configuration in which acoustic phenomena such as delay, masking, and reflection directly influence coordination. A more detailed discussion is provided in Section 2.3.

Within computer music, a pivotal contribution is Bisig, Neukom, and Flury's *Interactive Swarm Orchestra* (ISO), which articulated swarm intelligence as an Artificial Life–inspired paradigm [2]. However, most swarm-music systems remain primarily virtual: agents interact within simulated environments where proximity and coupling are abstracted, and sound functions largely as a symbolic or parametric variable.

By contrast, fully embodied musical swarms must operate within physical acoustic environments, where sound propagation, masking, reflection, and spatial heterogeneity fundamentally shape interaction. Prior work shows that local acoustic coupling can support decentralized musical coordination [16], but scaling exposes the limits of centralized timing, global clocks, and deterministic synchronization.

In this paper, we introduce *Sound Swarm*, an embodied, multi-agent system that functions as a self-organizing mesh synthesizer. Each node is a physically situated, autonomous sound-producing agent built on a minimal ESP32-S3 hardware platform. At the time of writing, the system has been prototyped in desktop-scale colonies (up to $N = 10$), with systematic data collection and benchmarking toward larger colonies ongoing. *Sound Swarm* operates as a *synthetic acoustic ecology*, in which musical form arises from the interaction of agent-level heuristics, the physical acoustic environment, and user-defined conditions. Our contributions are threefold: (1) a prototype method for collective spatial sense-making using calibration-derived relational tensors to infer relative topology without external tracking or absolute coordinates; (2) biologically inspired coordination behaviors that create, stabilize, and destabilize rhythmic and organizational regimes under masking and uncertainty; and (3) an observability and control layer (*swarmtool*) that supports emergent practice through telemetry and intervention.

To make this progression reviewable and reproducible, we develop topology inference in three stages: (a) synthetic feasibility validates the inference pipeline end-to-end, (b) tabletop prototyping confirms that the hardware yields usable pairwise RF/acoustic measurements, and (c) evaluation on real layouts benchmarks performance under room acoustics and multipath. This paper reports Stage (a) synthetic verification and an N=10 layout-split ablation study (all synthetic), and specifies the protocol and metrics used in the subsequent hardware and real-layout stages.

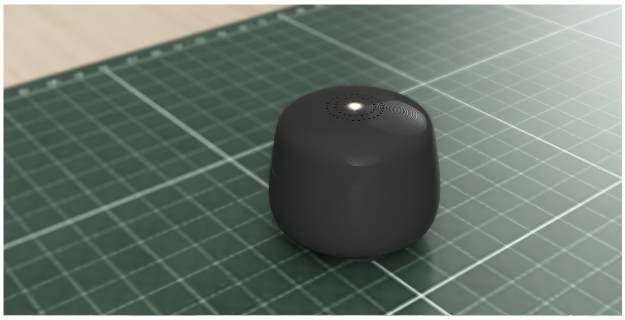

Figure 2: *Sound Swarm* node (v1 prototype): a self-contained ESP32-class microphone–speaker agent for spatial arrangement and environment-mediated interaction.

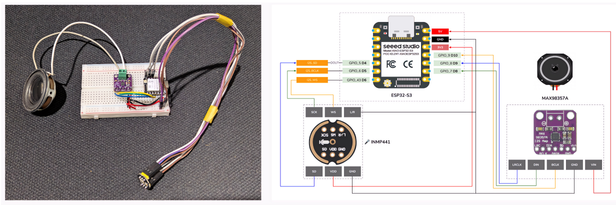

Figure 3: V1 node electronics and wiring, co-locating AMY synthesis, acoustic sensing, and low-rate coordination in a physical sound field.

2 Related Work

Research on swarm-based music systems spans algorithmic composition, interactive music systems, Artificial Life, and distributed audio. This section situates *Sound Swarm* within that landscape, highlighting how prior work addresses decentralized musical organization, and where limitations remain with respect to embodiment, environmental coupling, and autonomous agency.

2.1 Swarm Intelligence in Computer Music

Early applications of swarm intelligence to music were pioneered by Blackwell and colleagues, demonstrating that musical structure can emerge from local interaction rules rather than centralized control or score-based specification [3, 4]. These systems often map swarm dynamics (e.g., attraction/repulsion, local alignment) onto musical parameters, establishing a foundational principle for swarm music: coherence can arise without global coordination.

A seminal extension is Bisig, Neukom, and Flury's *Interactive Swarm Orchestra* (ISO), which articulated swarm intelligence as an Artificial Life–inspired paradigm for computer music [2]. ISO frames musical interaction as indirect: performers shape agent behavior through parameters and constraints rather than dictating events. This shift toward condition-setting is conceptually central, but ISO operates in virtual parameter spaces abstracted from the physical acoustic environment.

2.2 Distributed Audio and Large-Scale Playback

Distributed audio systems focus on engineering solutions for synchronized playback across spatially distributed nodes. Whitman's *Alles* demonstrates that large arrays of networked speakers can be coordinated using UDP multicast over Wi-Fi, enabling distributed synthesis and polyphony at scale [25]. Large-scale installations such as *Bloom* deploy hundreds to thousands of distributed audio-visual devices using centralized sequencing and carefully managed synchronization strategies [9]. These systems address deployment and scale, but generally retain a top-down orchestration model: nodes act as rendering endpoints for centrally defined behavior, and the environment is treated primarily as a transmission medium rather than an active participant in musical organization.

2.3 Embodiment and Physical Swarm Systems

As introduced earlier, we use *embodiment* to refer to systems in which sound-producing agents are physically co-present and interact through real acoustic propagation, such that the environment participates in shaping interaction rather than acting as a neutral transmission channel. In this sense, embodiment includes not only the presence of physical sound sources, but also the spatial and topological relationships between agents and the shared acoustic field through which they perceive and influence one another.

Within this framing, embodiment is closely tied to *condition-setting* as a compositional approach. In embodied swarm systems, conditions are not purely abstract parameters, but are expressed through spatial arrangement and environment-mediated interaction. This relationship is developed further in Section 3.

Prior work provides important reference points but also highlights a gap. For example, Krzyżaniak's *Dr. Squiggles* demonstrates autonomous rhythmic robots coordinating via local acoustic coupling using predictive beat tracking and exhibiting stable yet perturbable equilibria [16]. However, contact microphones intentionally isolate interaction via solid-body transmission, sidestepping masking and spatial heterogeneity that characterize open-air sound fields.

Sound Swarm aims to shift coordination back into the acoustic field itself, relating our rhythmic tests to frog-chorus models of masking avoidance and phase organization [1, 13]. In this context, coordination emerges through a shared acoustic medium in which agents both generate and perceive sound within the same spatial field.

Sound art installations offer a complementary direction. Mihara and Saita's *Moids 2.2.0* comprises large populations of autonomous electroacoustic agents that respond both to ambient sound and to their own output, framing a decentrally organized "alternative nature" [20]. While such systems foreground acoustic emergence, they do not explicitly support shaping interaction through spatial configuration or inferred relational structure.

Taken together, these approaches suggest a gap: while embodied and acoustically responsive systems exist, fewer treat the acoustic field as a *shared medium of agency*, in which spatial arrangement and topology directly shape interaction and can be configured through condition-setting.

Sound Swarm is positioned as a step toward this direction. Rather than resolving embodiment in full, this work explores a configuration in which agents generate and perceive sound within a shared acoustic field, and where rearrangement alters the structure of interaction itself. In this sense, the contribution lies in instantiating an embodied configuration in which spatial arrangement, acoustic propagation, and agent behavior form a coupled interaction loop.

2.4 Human–Swarm Interactive Music Systems

Lucas and Glette propose a framework for Human–Swarm Interactive Music Systems (HS-IMS), covering design principles, technologies, and evaluation approaches across virtual, hybrid, and physical implementations [17]. Their later platform work emphasizes practical tooling for experimentation and evaluation in HS-IMS settings [18]. These contributions clarify a persistent gap: many systems emphasize either virtual swarm dynamics or physical distribution, but fewer integrate autonomous agency, embodied interaction, and environmental coupling within a single scalable system. *Sound Swarm* builds on this line by unifying embodied sound production, distributed sensing, and topology inference as a continuous interaction loop.

2.5 Ecosystemic and Biologically Inspired Coordination

Di Scipio's ecosystemic signal processing reframes sound as both signal and environment, emphasizing feedback, instability, and environmental mediation as compositional materials rather than artifacts to be controlled [7]. Behavioral objects reconceptualize musical systems as entities with internal agency that negotiate interaction over time [5].

Biological coordination mechanisms further inform swarm-based design. Stigmergy, articulated by Grassé as indirect coordination via environmental modification, has been extended to computational multi-agent systems [12, 22]. Studies of collective decision-making in honeybee swarms illustrate how large groups achieve coherence through local interaction and role differentiation [24]. Frog chorus models demonstrate how temporal desynchronization and anti-phase behavior mitigate masking in dense sonic environments [1, 13].

Together, this work establishes the foundation for *Sound Swarm*: a synthetic acoustic ecology that adds explicit spatial condition-setting to embodied swarms, linking rearrangement to inferred neighborhood structure and topology-conditioned behavior within a deployable embedded architecture.

3 Problem Framing: Condition-Setting in Embodied Acoustic Swarms

The preceding sections establish swarm-based approaches as alternatives to event-driven composition and show how embodied, physically distributed sound systems introduce spatial and acoustic constraints. When interaction occurs through a shared acoustic field, arrangement, propagation, and environmental coupling shape behavior directly, making event-level specification increasingly unstable.

We therefore frame *condition-setting* as a principled strategy for organizing sound in embodied, decentralized systems, triangulating engineering limits, compositional theory, and biological coordination.

3.1 Scale displaces events in favor of populations

At sufficient scale, musical organization shifts from individual events to collective behavior. Distributed installations show that event-level control becomes perceptually unreliable as spatial dispersion, latency, and jitter increase; form must often be expressed statistically (e.g., densities, constraints, and probabilities) rather than through deterministic scheduling. This aligns with swarm-music practice: ISO formalizes the idea that complex musical systems are better navigated by designing intermediaries rather than specifying outcomes [2]. This perspective also resonates with earlier formulations of interactive music systems, in which the composer defines processes and interaction behaviors rather than fixed compositional outcomes [6]. Scale therefore motivates population-level control both practically and compositionally.

3.2 Embodiment constitutes a categorical shift

HS-IMS frameworks distinguish virtual, hybrid, and physical swarms, noting that many implementations remain virtual, where proximity and interaction are abstracted [17]. Fully physical systems—where agents share space with listeners—introduce qualitatively different constraints: spatial arrangement, occlusion, and acoustic propagation are not parameters but causal forces.

As a result, spatial configuration becomes a compositional variable rather than a static substrate. In this sense, embodiment does not simply add physical presence, but transforms the problem of musical control, requiring strategies that operate on interaction conditions rather than predefined events.

3.3 Geometry replaces temporal scoring

Biological acoustic systems provide concrete demonstrations of this shift. Models of Japanese tree frog choruses show that collective rhythmic behavior can be strongly shaped by boundary conditions and spatial arrangement: the same local rules yield distinct global phase patterns across different geometries [1].

This suggests that geometry can function as a score: arranging agents in space can induce stability, alternation, or restless variation without centralized timing control.

3.4 Sound operates as both signal and constraint

Frog chorus studies also show that coordination is driven by masking: individuals adopt temporal spacing to remain audible in a shared field [1, 13]. In dense environments, reflection and reverberation prevent sound from acting as a transparent one-to-one communication channel; interference becomes an organizing force. Here, *noise* refers not to arbitrary unwanted sound or statistical measurement error, but to non-target acoustic energy in the shared field—including masking, reverberant overlap, reflections, and bleed between sources—that degrades the signal available to other agents. Rather than treating such noise as something to suppress, *Sound Swarm* seeks to embrace it as part of the perceptual medium through which agents coordinate and adapt [7].

3.5 Environment as shared memory and mediator

HS-IMS frameworks identify stigmergy—indirect coordination via environmental modification—as a core design principle when direct control of large populations is infeasible [12, 22].

Sound Swarm extends this principle by treating the acoustic field itself as the stigmergic medium: reverberation, spectral occupation, and temporal overlap become persistent traces that bias future behavior. In this formulation, the environment functions not only as a transmission medium, but as a distributed memory that mediates interaction across agents.

Positioning Summary. Engineering studies motivate statistical control at scale; swarm-music theory motivates intermediary, condition-based systems; HS-IMS frameworks formalize decentralization, embodiment, and stigmergy; and biological acoustic systems show that spatial constraints can generate stable or dynamic rhythmic organization through local interaction.

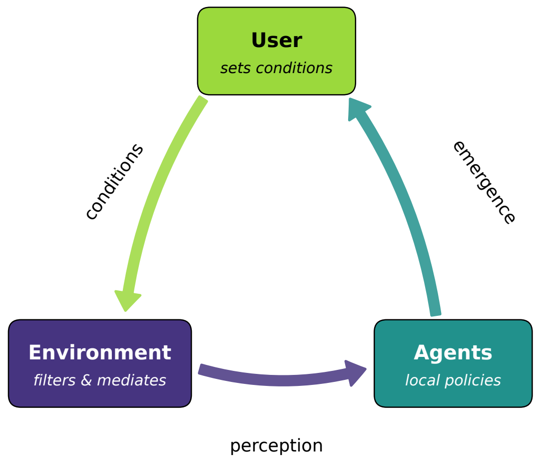

Figure 4: Triadic Interaction Loop. Extending interactive-system models (e.g., Chadabe), condition-setting links user-defined arrangements, environmental mediation through propagation, and agent-level local policies. Musical organization emerges through recursive feedback.

Sound Swarm operationalizes condition-setting in the physical domain by treating geometry, perception, and environment as primary organizing forces, extending these ideas into a setting where spatial arrangement and acoustic coupling directly structure interaction.

4 Design Framing: The Synthetic Ecology

We frame *Sound Swarm* not as an instrument ensemble or centralized interactive system, but as a *synthetic ecology*: a population of embodied agents whose sonic organization emerges from local rules coupled through the physical acoustic environment. Rather than specifying musical events, the composer establishes *conditions*—interaction policies, spatial constraints, and time-scales—under which patterned behavior can self-organize.

This framing draws from biological self-organization, where global structure arises from local interaction without centralized control [11]. It extends prior swarm-music approaches [2, 4] by locating emergence in *environmentally mediated acoustic coupling* rather than abstract simulation alone.

Sound-art precedents further motivate this view. Mihara and Saita's *Moids* pursues a decentrally organized "alternative nature" through large populations of microphone–microcontroller–speaker circuits that respond to ambient and self-produced sound [20]. *Sound Swarm* builds on this lineage while emphasizing *condition-setting*, explicit sensing, coordination, and topology-aware control for embodied swarm performance.

4.1 The Triadic Interaction Loop

Emergent soundscapes arise from a closed feedback loop between three components:

(1) **User — Condition Setter.** The user configures density, spatial topology, behavioral thresholds, and role constraints as boundary conditions rather than event-level instructions.
(2) **Environment — Acoustic Medium.** The physical environment filters, delays, and redistributes sound, functioning as a stigmergic substrate that stores and transforms traces [12, 22].
(3) **Agents — Local Policy Executors.** Each node responds to local perception and inferred relational structure; no agent has global knowledge, and organization emerges via repeated local adaptation [16].

4.2 Design Principles for Condition-Setting

From this ecological framing, we derive five design principles.

P1. Compose interaction rules, not sounds. Musical structure arises from local policies rather than explicit events [4].

P2. Compose spatial conditions (topology as control). Rearrangement changes coupling and neighborhood structure, shifting system behavior.

P3. Compose time-scales, not synchronization. Agents operate without a shared clock; equilibria and instabilities emerge from local timing adaptation [1, 16].

P4. Compose stigmergic adaptation. Acoustic traces bias future behavior, shifting authorship toward environmental mediation [11, 12].

P5. Compose tendencies, not outcomes. Thresholds and probabilistic gains create bifurcations in global structure under small parameter changes [11].

5 System Architecture

To implement *Sound Swarm* as a synthetic acoustic ecology, we use a distributed hardware–software architecture in which each node is an autonomous embedded synthesizer agent. The system extends modular synthesis by relocating patching from cable-based signal routing into embodied interaction: each node generates sound locally, listens to the acoustic field, and adapts its synthesis state from what it perceives. Musical structure is therefore shaped by spatial arrangement and environment-mediated coupling.

Low-cost microcontrollers now make it feasible to co-locate real-time synthesis, sensing, and coordination on each node. *AMY* provides the embedded sound engine [26], allowing nodes to function as independent voices rather than playback endpoints.

5.1 Synth Node Hardware (Embodied Voice Unit)

Each node integrates (i) a MEMS microphone for acoustic perception, (ii) an embedded processor (ESP32-S3 class) for synthesis and control, (iii) an amplifier and speaker for sound output, (iv) a battery for untethered deployment, and (v) a status LED used for calibration cues and runtime state feedback. This design makes each node a self-contained "voice" whose placement in space is itself a compositional act: rearranging nodes changes local sensing, coupling, and the effective interaction topology.

5.2 Runtime Architecture: Expression (Fast Sound) and Behavior (Slow Cognition)

Embedded musical agents must satisfy two competing requirements: stable audio-rate synthesis and continuous sensing, communication, and decision-making under tight compute and memory constraints [19]. We address this by separating audio-rate synthesis from low-rate control through a **decoupled dual-core architecture** on the ESP32-S3.

The **Expression layer** runs at high priority and is dedicated to real-time audio synthesis. We employ *AMY* as the embedded sound engine [26], allowing each node to render sound locally with predictable timing. Expression exposes a set of smoothly varying parameters (e.g., pitch targets, amplitude envelopes, timbral controls) that are modulated by higher-level logic without compromising audio stability.

The **Behavior layer** runs on the second core at a slower control rate (typically 30–50 Hz). It performs perception updates, network messaging, rule evaluation, and coordination logic, producing control signals that are passed to Expression and interpolated at audio rate. This split allows nodes to adapt (e.g., respond to masking, participate in role election, or react to inferred topology) without interrupting their sound output.

5.3 Calibration Roles: Queen and Drones

To coordinate time-critical calibration exchanges without external infrastructure, one node is temporarily elected as a coordinator ("queen") that schedules probe turns and aggregates pairwise measurements; all other nodes act as drones. The election mechanism is intentionally lightweight and reversible, borrowing the idea of transient role differentiation from honeybee swarms [24]. These roles exist to support sensing and calibration rather than to impose a musical hierarchy; once a Ghost Map is computed, control returns to local agent policies (Section 6).

5.4 On-Device Perception and Calibration Signaling

Each node monitors its acoustic environment via an onboard microphone and computes lightweight features (e.g., RMS amplitude, ZCR, coarse band-energy measures) to support local state transitions and masking-aware behavior. Perception also supports collective inference: a calibration ritual can be triggered on demand (e.g., after rearrangement) to populate pairwise RF/acoustic measurements. To reduce perceptual intrusion, probe signals can occupy a reserved narrow frequency region that musical synthesis avoids via spectral notching; higher-frequency and ultrasonic probing remain a design direction, limited by transducer bandwidth in the current prototype.

5.5 Observability and Control: *swarmtool*

Because emergent behavior in decentralized systems cannot be fully predicted from local rules alone, practical experimentation and performance require observability and carefully scoped intervention. We therefore treat *swarmtool* as a *condition-setting console* rather than a centralized controller: it makes collective state legible, enables rapid iteration on boundary conditions, and supports calibration rituals (cf. tooling needs in HS-IMS contexts [18]). Nodes remain autonomous when the tool is absent.

swarmtool provides:

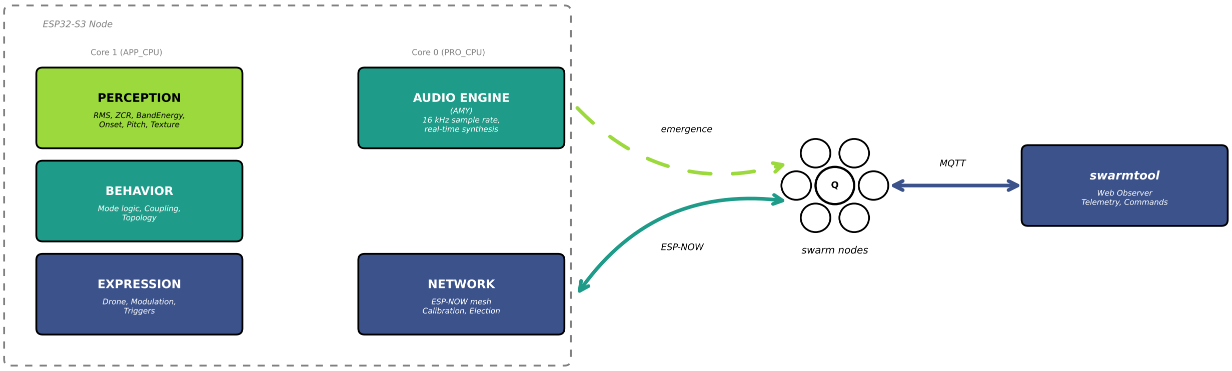

Figure 5: Software architecture of a *Sound Swarm* node and swarm interface. Core 0 runs AMY audio at 16 kHz; Core 1 runs Perception–Behavior–Expression and networking at ∼10–50 Hz. Nodes self-organize over ESP-NOW, with task-scoped calibration coordination and MQTT telemetry/commands to *swarmtool*.

Figure 6: Exploded *Sound Swarm* node showing the MEMS microphone, status LED, ESP32-S3 MCU, audio amplifier/DAC, LiPo battery, speaker, and enclosure.

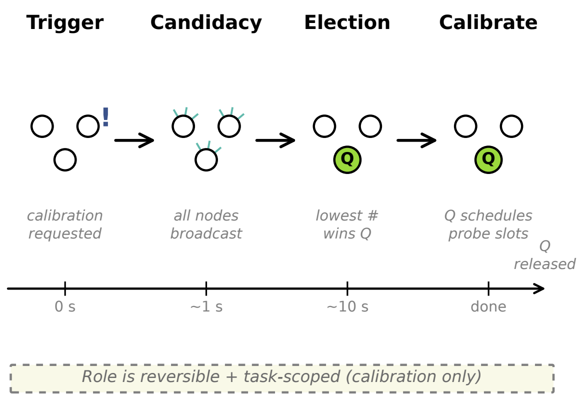

Figure 7: Honeybee-inspired temporary role election for calibration scheduling. Nodes broadcast candidacy signals; a short-epoch coordinator ("queen") schedules probe turns and can relinquish leadership.

- **Topology-aware observability:** telemetry plus Ghost Map visualization to interpret neighborhood structure during rearrangement and calibration.
- **Condition-setting controls:** broadcast parameter updates that shape local policies (e.g., coupling gains, inhibition thresholds, role-election parameters) without event scheduling.
- **Rapid deployment:** bulk distribution of synthesis profiles (AMY parameter sets) and thresholds via a coordinator gateway, reducing iteration time compared to per-node reflashing.

Roadmap note. To foreground the ecological interaction claim, we describe two implemented bio-inspired behavior exemplars next (Section 6), then detail the Perception mechanism (collective topology inference) that parameterizes neighborhood-conditioned interaction (Section 7).

6 Behavior: Biologically Inspired Coordination and Emergent Musical Form

The Behavior layer transforms perception into adaptive musical action using biologically grounded heuristics under uncertainty [5, 17]. In *Sound Swarm*, behavior is expressed as local policy: agents adjust timing, influence, and participation based on immediate acoustic perception and (when available) inferred relational structure. We foreground two implemented exemplars—role differentiation and inhibitory rhythmic coupling—to make the "synthetic ecology" framing concrete before introducing the topology inference mechanism that supports spatial condition-setting (Section 7).

6.1 Role Election and Task-Scoped Hierarchy

Honeybee-inspired threshold dynamics motivate decentralized role differentiation [24]. *Sound Swarm* implements temporary role election (Figure 7) to select a coordinator for time-critical calibration scheduling. The role is reversible and intentionally scoped: it exists to structure sensing/communication rituals rather than to impose a musical hierarchy. In future work, the same role mechanism can be extended to musical leadership behaviors (e.g.,

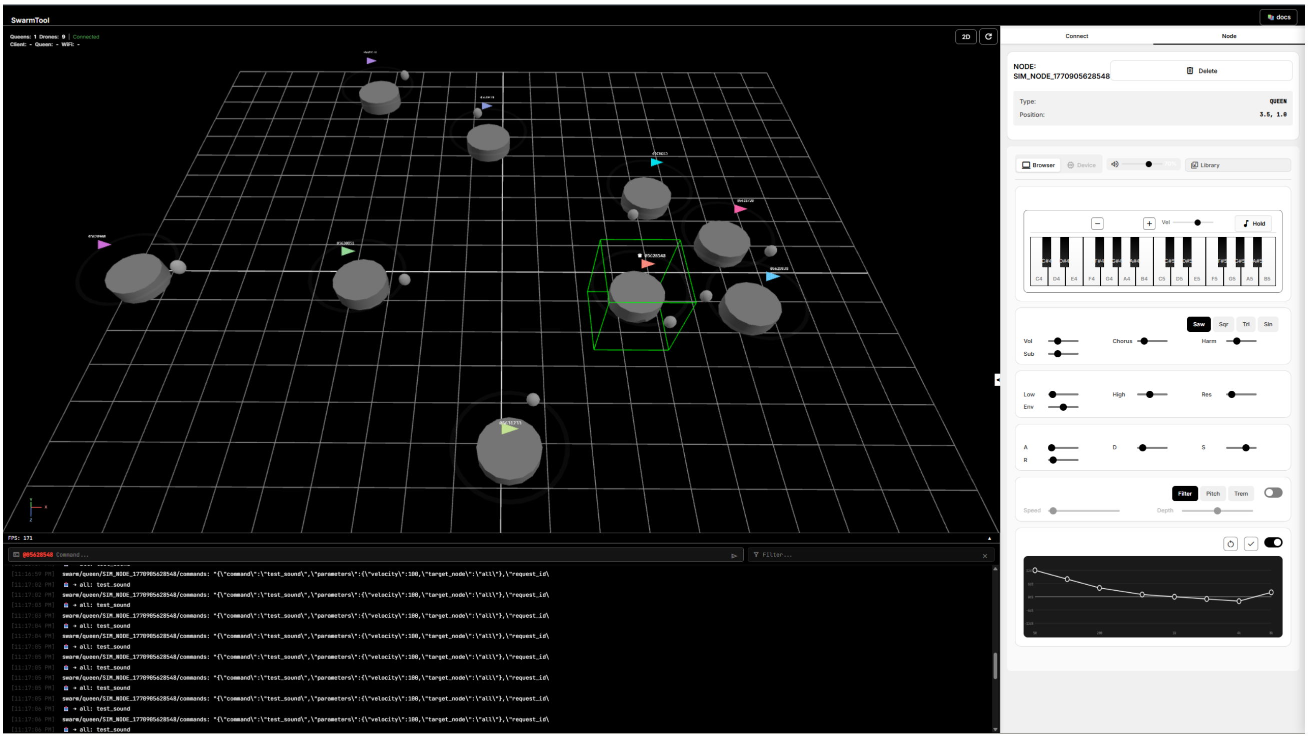

Figure 8: *swarmtool* interface for telemetry, topology visualization, condition-setting, voice-profile definition, and rapid swarm configuration without per-node reflashing.

initiators or "tempo-setters") while remaining decentralized and re-electable.

6.2 Rhythmic Coordination and Equilibrium Dynamics

Frog choruses provide a model of masking avoidance and phase organization without global clocks [1, 13]. Agents implement analogous local timing adaptation: rather than synchronizing, they repel overlap to maintain audibility in a shared acoustic field.

Frog-inspired inhibitory coupling (implemented policy). Each node maintains an internal calling phase $\theta_i \in [0, 2\pi)$ with natural frequency ω_i and renders a short call event when θ_i crosses 2π. Neighbor calls are detected locally from microphone energy features; detected events induce a bounded inhibitory phase update that reduces overlap:

$$\frac{d\theta_i}{dt} = \omega_i + \sum_{j \in \mathcal{N}_i} \kappa_{ij}\, g\big(\mathrm{wrap}(\theta_j - \theta_i)\big), \tag{1}$$

where $\mathcal{N}_i$ denotes acoustically salient neighbors, κ_{ij} is a coupling gain, and $g(\cdot)$ repels overlap (e.g., $g(\Delta) = -\sin(\Delta)$). In firmware, we implement a discrete-time bounded phase shift for stability under detection jitter.

Bench-top test configuration. Pilot phase experiments were conducted on a tabletop using open-air acoustic coupling between self-contained nodes (Figure 10). Nodes interacted through local microphone sensing only (no shared clock or wired synchronization), and the same firmware policy was used unchanged in both the $N = 2$ and $N = 3$ trials.

Preliminary hardware validation (pilot). On v1 hardware, two nodes can converge to a stable alternation (anti-phase) regime using only local acoustic perception. In pilot three-node trials, phase-spaced organization consistent with a tri-phase tendency can also be elicited. Figure 9 summarizes representative trajectories (smoothed for clarity): the two-node condition approaches $\Delta\theta \to \pi$, while the three-node pilot organizes toward approximately equal spacing ($\Delta\theta \approx 2\pi/3$), consistent with inhibitory phase-oscillation models of Japanese tree frog calling behavior [14]. Systematic evaluation across densities and room conditions is ongoing.

6.3 Acoustic Stigmergy and Distributed Sense-Making

Stigmergy enables coordination through a shared medium [12, 22]. In *Sound Swarm*, we treat the acoustic field as a stigmergic medium: spectral occupation, reverberant persistence, masking, reflections, and temporal overlap can leave short-lived traces that bias subsequent behavior. These traces do not encode explicit symbolic state, but they change what nearby agents are able to perceive and therefore influence local timing, inhibition, and response. Coordination is therefore partly environmentally mediated rather than purely message-based, consistent with ecosystemic views of sound as both signal and environment [7].

6.4 Topology-Conditioned Coupling as Condition-Setting

As swarms scale, not all neighbors should contribute equally. We therefore treat coupling gains and thresholds as *conditions*

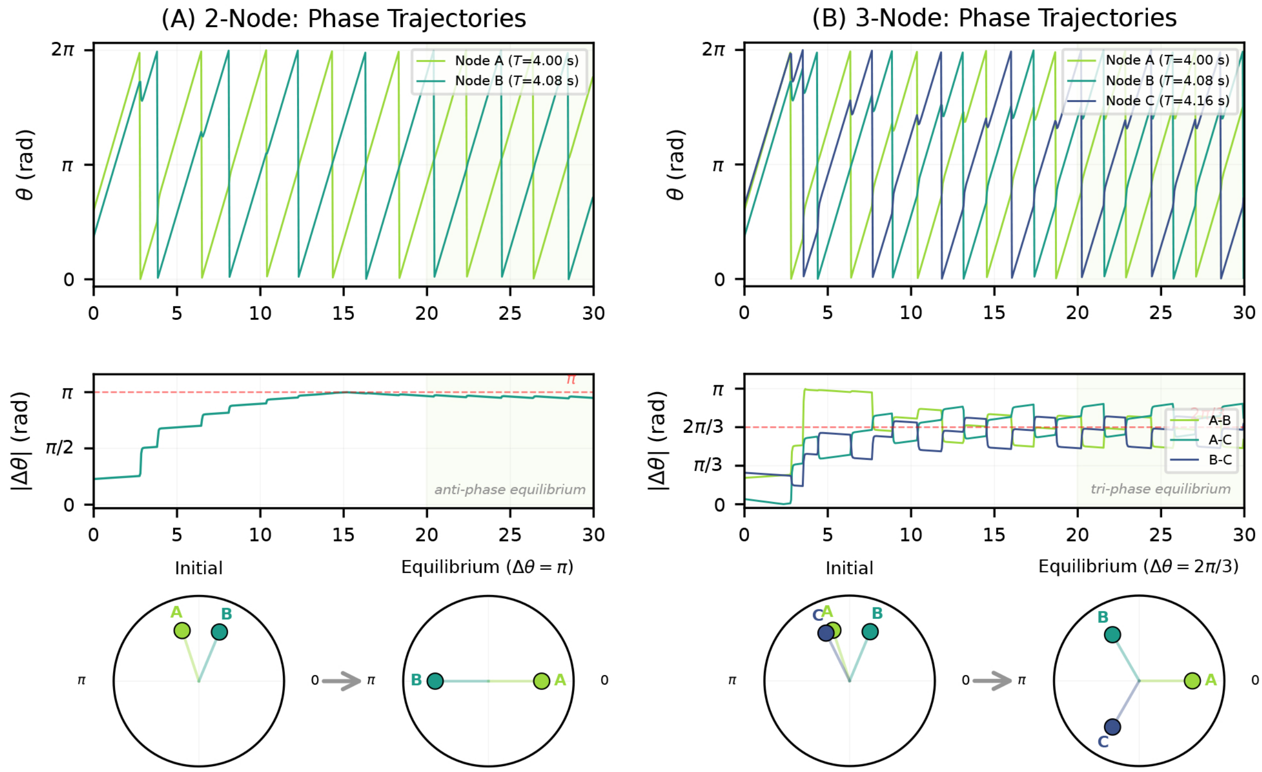

Figure 9: Pilot oscillator phase experiments on v1 nodes, smoothed from telemetry. (A) Two-node inhibitory coupling converges to anti-phase alternation ($\Delta\theta \to \pi$). (B) Three-node trials approach tri-phase spacing ($\Delta\theta \approx 2\pi/3$), consistent with Japanese tree frog inhibitory phase-oscillation models [14].

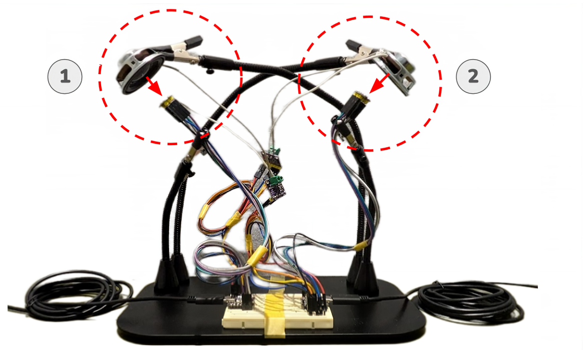

Figure 10: Bench-top oscillator phase-coupling test with two open-air *Sound Swarm* nodes executing the inhibitory phase-update policy in Eq. 1, without shared clock or wired synchronization.

derived from perceived proximity and local density. In the current prototype, proximity can be estimated directly from pairwise sensing, and (when available) from the inferred Ghost Map (Section 7). This closes the interaction loop required for spatial condition-setting: *rearrangement* changes relational perception, which re-parameterizes local policies. Composition is thus performed as spatial lutherie rather than sequenced instruction [5, 8].

7 Perception: Collective Spatial Sense-Making

The behavior exemplars above rely on locality: agents must determine *who is near whom* to modulate coupling and density-sensitive thresholds. For this to function as an interaction method, the swarm must infer relational neighborhood structure from situated sensing, without external tracking, absolute coordinates, or a global clock. We implement a relational perception pipeline that converts calibration-derived pairwise measurements into a topology-aware latent map (a "Ghost Map") used to parameterize neighborhood-conditioned behaviors.

Staged development. We develop topology inference through a simulation-first progression: (a) *synthetic feasibility* verifies the end-to-end mapping under controlled generative models; (b) *tabletop prototyping* validates measurement quality on hardware; and (c) *room evaluation* benchmarks the full loop under multipath, masking, and occlusion. This paper reports Stage (a) results and specifies the protocol and metrics used for Stages (b)/(c).

7.1 Online Inference: Arrangement → Tensor → Topology

When the user rearranges nodes, *Sound Swarm* runs a calibration ritual to estimate the current *relational structure* of the swarm. A designated coordinator schedules a round-robin protocol: each node emits a standardized probe while all other nodes listen and record observations. Measurements are indexed by ordered pairs ($i \to j$), preserving real-world asymmetries (e.g., orientation, occlusion, room response).

In the current prototype, two modalities are used:

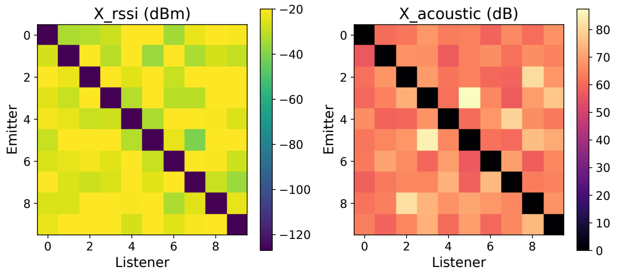

Figure 11: Relational tensor slices for topology inference: RF signal strength (RSSI) and received acoustic level. Rows are emitters, columns receivers; diagonal self-interactions are omitted.

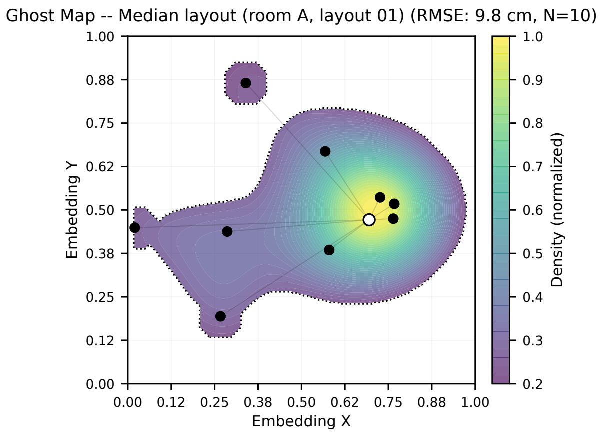

Figure 12: Ghost Map from the $N = 10$ Stage (a) synthetic evaluation, median case; positional RMSE 10.2 cm. Pairwise RF/acoustic coupling yields a relative 2D embedding; color indicates local density, dotted boundary the inferred footprint, and white the elected coordinator.

- **RF signal strength (RSSI):** received signal strength from ESP-NOW metadata (coarse and noisy indoors).
- **Acoustic received level:** received acoustic energy during a short probe tone/chirp (optionally placed in a spectrally notched channel).

Pairwise measurements are assembled into a relational tensor

$$\mathbf{X} \in \mathbb{R}^{N \times N \times F}, \qquad F = 2, \tag{2}$$

where each entry $\mathbf{X}[i, j, :]$ encodes how receiver j perceives emitter i across modalities.

Online topology inference converts $\mathbf{X}$ into a relative embedding (the Ghost Map) used primarily for neighborhood structure rather than absolute coordinates. In the deployed pipeline (Section 7.3), a shared per-edge neural model maps each $\mathbf{x}_{ij}$ to an estimated pairwise distance $\hat{d}_{ij}$ to form $\hat{\mathbf{D}}$; classical multidimensional scaling (MDS) then produces a 2D embedding. Changes in physical arrangement therefore propagate as: arrangement $\rightarrow \mathbf{X} \rightarrow$ Ghost Map $\rightarrow$ topology-conditioned coupling and synthesis constraints.

7.2 Offline Training and Evaluation Protocol

Offline training defines supervision and metrics. Ground-truth planar layouts $\mathbf{P}^{gt} \in \mathbb{R}^{N \times 2}$ are collected in a controlled tabletop

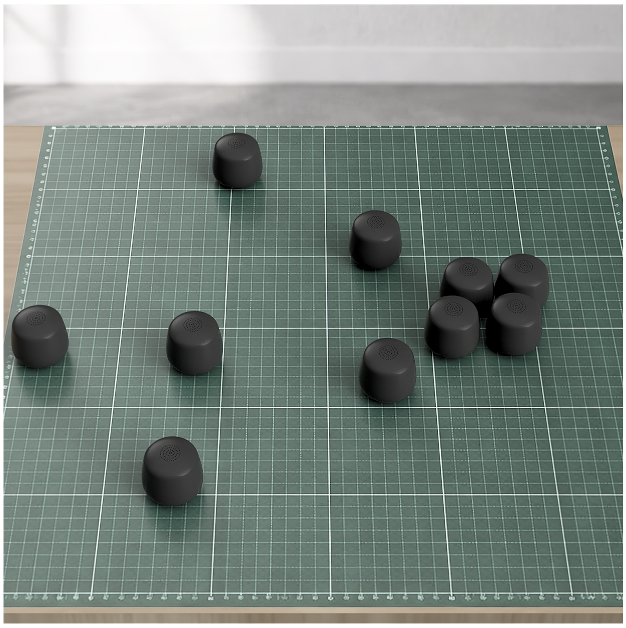

Figure 13: Offline training data collection. *Sound Swarm* nodes are arranged on a 1 m × 1 m surface; *swarmtool* records node IDs and ground-truth coordinates paired with calibration-derived relational tensors for distance regression.

setup (100 cm × 100 cm), either by measured grid placement or overhead tracking. From $\mathbf{P}^{gt}$ we compute pairwise distances

$$d_{ij}^{gt} = \left\| \mathbf{p}_i^{gt} - \mathbf{p}_j^{gt} \right\|_2, \tag{3}$$

and train a distance regressor on tuples $(\mathbf{x}_{ij}, d_{ij}^{gt})$ using per-edge MSE.

At evaluation time, predicted distances form $\hat{\mathbf{D}}$, MDS yields $\hat{\mathbf{P}}$, and two topology-quality metrics are reported: **positional RMSE** after Procrustes alignment (rigid-transform ambiguity) and k**NN accuracy** (typically $k = 2$) as a neighborhood-utility measure for downstream coupling.

Development note. At the time of writing, systematic labeled datasets across environments are still in progress. We therefore present topology inference primarily as an implemented interaction mechanism (arrangement $\rightarrow \mathbf{X} \rightarrow$ Ghost Map) and report Stage (a) synthetic experiments to unit-test the pipeline and explore stress conditions; these results are not treated as evidence of real-room performance.

7.3 Model: Shared Edge Encoder + MDS

To remain deployable on embedded hardware while operating on relational observations, we use a two-stage pipeline: (1) a shared *edge encoder* maps each pairwise measurement vector $\mathbf{x}_{ij}$ to an estimated distance,

$$\hat{d}_{ij} = h_\theta(\mathbf{x}_{ij}), \tag{4}$$

and (2) classical MDS converts the resulting distance matrix $\hat{\mathbf{D}}$ into a 2D relative embedding $\hat{\mathbf{P}} \in \mathbb{R}^{N \times 2}$.

The edge encoder h_θ is implemented as a small MLP with shared weights across all ordered pairs. In the current prototype, $\mathbf{x}_{ij}$ has F=2 features (RSSI and received acoustic level), and we use a 2⟶16⟶8⟶1 architecture (ReLU activations) with a sigmoid-scaled output to $[0, d_{\max}]$. This model has 193 parameters (0.75 KB in float32), enabling on-device inference on ESP32-class hardware. The neural component is analogous to an *edge model* in neural relational inference frameworks [15], while the global embedding is enforced by the deterministic MDS step.

Embedding via Classical MDS. This topology inference step is a distance-geometry problem: given noisy pairwise distance estimates, recover a planar configuration. Classical (metric) MDS is theoretically well-matched to this setting: when $\hat{\mathbf{D}}$ is (approximately) a Euclidean distance matrix, the eigendecomposition of the doubly-centered squared-distance structure recovers the generating coordinates exactly up to rigid transforms; under measurement noise, retaining the top k=2 components yields a principled low-distortion embedding (a best rank-2 approximation of the induced Gram structure) without adding learnable parameters or hyperparameter tuning. Because pairwise distances are invariant to translation, rotation, and reflection, we evaluate embeddings after Procrustes alignment.

Directed estimates are symmetrized (e.g., $\hat{d}_{ij}$ and $\hat{d}_{ji}$ averaged) and the diagonal is set to zero. Downstream behaviors consume topology through relational queries (neighbors, density, cluster structure) rather than absolute coordinates.

Scaling considerations. The presented architecture is deliberately minimal to match embedded constraints. As swarm sizes increase, $O(N^2)$ measurement cost and partial observability motivate sparse sensing and lightweight message-passing extensions; we treat these as future work and avoid performance claims absent empirical evaluation.

8 Evaluation and Emergent Outcomes

At the time of writing, empirical evaluation includes prototype colonies up to $N = 10$; broader benchmarking across environments and larger colonies is ongoing. Accordingly, this section (i) specifies the evaluation protocol and metrics used to test the Perception pipeline and (ii) reports Stage (a) synthetic results that unit-test topology inference. These synthetic experiments are used for verification and hypothesis development and are *not* treated as evidence of real-room performance. Preliminary behavior demonstrations on hardware (role election and inhibitory coupling) are summarized in Section 6.

8.1 Experiment: Topology Inference — Synthetic Verification and N=10 Ablation

We evaluate topology inference under a staged synthetic progression. Across studies, we report three metrics: (i) **distance RMSE** (per-edge distance error), (ii) **position RMSE** (Procrustes-aligned error after MDS), and (iii) k**NN accuracy** (with $k = 2$) as a topology-utility metric for neighborhood-conditioned behavior.

8.1.1 Phase 0: Synthetic end-to-end verification (Classical MDS). We first unit-test the full pipeline on synthetic layouts in a 100 cm × 100 cm workspace ($N = 6$), simulating RSSI and received acoustic level as log-distance laws in dB with additive noise and practical clipping/quantization for RSSI:

$$\mathrm{RSSI}_{ij} = A_{\mathrm{rssi}} - 10 n_{\mathrm{rssi}} \log_{10}\left(\frac{d_{ij}}{d_{\mathrm{ref}}}\right) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2_{\mathrm{rssi}}), \tag{5}$$

$$L_{ij} = A_{\mathrm{ac}} - 20 \log_{10}\left(\frac{d_{ij}}{d_{\mathrm{ref}}}\right) + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2_{\mathrm{ac}}). \tag{6}$$

This phase is used to verify that plausible propagation plus noise yields recoverable *coarse topology* under the evaluation protocol (MDS + Procrustes + kNN), not to model real rooms.

Table 1: Phase 0 MDS reconstruction error (synthetic, N = 10 in a 100 cm workspace).

Method	Mean RMSE (cm)	Std RMSE (cm)	Mean (%)
RSSI-only	22.48	11.73	22.5%
Acoustic-only	24.05	14.69	24.1%
Combined (α=0.5)	21.38	14.82	21.4%

Table 2: Synthetic ($N = 10$) baseline vs. learned edge encoder (Phases 1–2).

Variant	Dist. RMSE (cm)	Pos. RMSE (cm)	kNN Acc. (k=2)
Calibrated baseline	18.91	42.67 ± 20.00	71.7%
Learned edge MLP	11.83	33.91 ± 18.42	82.5%

Table 3: N=10 ablation results on the validation set (12 runs; synthetic).

Variant	Dist. RMSE (cm)	Pos. RMSE (cm)	kNN Acc. (k=2)
RSSI-only	15.15 ± 2.11	20.52 ± 15.98	67.5% ± 13.8%
Combined	**10.03 ± 1.78**	22.57 ± 17.00	**77.9% ± 8.8%**
Acoustic-only	11.62 ± 2.04	22.57 ± 18.01	71.7% ± 12.6%

8.1.2 Phases 1–2: Baseline vs. learned edge encoder. We next compare a calibrated analytic baseline (invert fitted propagation laws, fuse modalities, then MDS) against the learned edge encoder (Section 7.3). The goal is improved *topology utility* after MDS under the same synthetic conditions ($N = 10$).

8.1.3 Paper Evaluation: N=10 layout-split validation and modality ablation. We apply the same learned edge encoder + MDS pipeline at the prototype-relevant scale ($N = 10$) using a layout-split protocol: all measurements from a subset of layouts are held out for validation. We simulate 60 runs (2 synthetic "rooms" × 10 layouts/room × 3 repeats/layout), split by layout (48 train / 12 validation), and train the same 193-parameter distance MLP. We evaluate three ablations: RSSI-only, acoustic-only, and combined.

Interpretation. Three observations are most relevant for interaction design. (1) **Multi-modal sensing improves distance estimation:** the combined model achieves 10.03 ± 1.78 cm distance RMSE, a 34% reduction relative to RSSI-only. (2) **Position RMSE exhibits high variance:** MDS is sensitive to noise and geometric ambiguity, leading to large spread in positional error even when distance RMSE improves. (3) **Neighborhood accuracy is comparatively stable:** the combined model achieves 77.9% ± 8.8% kNN accuracy ($k = 2$), aligning with our claim that topology is primarily used to parameterize local coupling and density rather than to provide metrically precise coordinates.

8.2 Experiment: Minimal Topology→Synthesis Demonstration (Illustrative)

To keep topology-conditioned behavior claims concrete without overstating empirical coverage, we specify a minimal mapping from one topology scalar to one synthesis parameter and illustrate its effect. We choose **graph density** and map it to **harmonic richness** (number of additive partials).

Topology scalar. Given inferred positions $\hat{\mathbf{P}}$, we form an undirected graph by connecting node pairs within a fixed radius

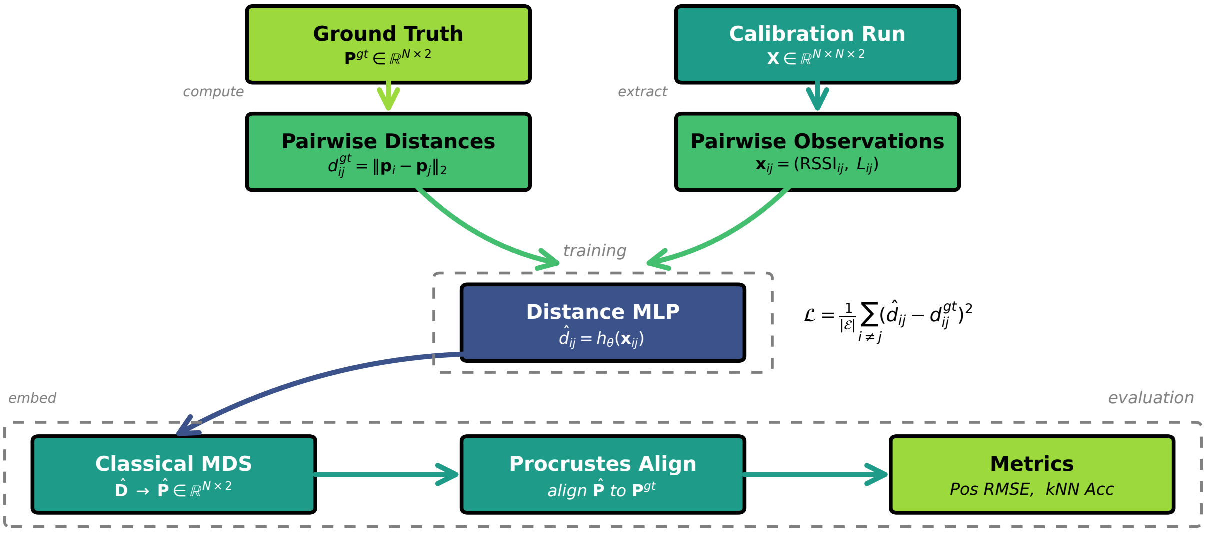

Figure 14: Offline training and evaluation loop. Pairwise observations train the distance model with per-edge MSE; predicted distances are embedded with MDS and evaluated by Procrustes-aligned RMSE and neighborhood accuracy.

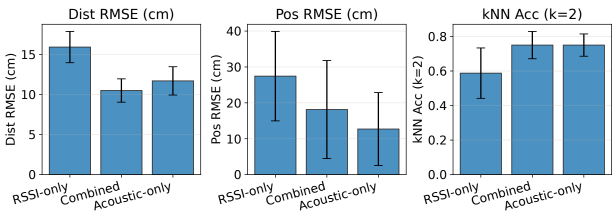

Figure 15: $N = 10$ synthetic ablation summary. Bars show mean ± std across 12 held-out runs for distance RMSE, position RMSE, and kNN accuracy ($k = 2$).

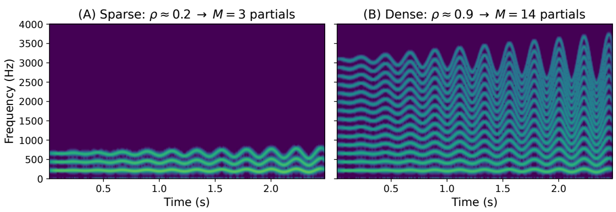

Figure 16: Minimal topology→synthesis demonstration. Sparse topology ($\rho \approx 0.2$) yields $M = 3$ additive partials; dense topology ($\rho \approx 0.9$) yields $M = 14$, producing a clear timbral contrast at 16 kHz.

τ and compute density

$$\rho = \frac{|E|}{\binom{N}{2}}. \tag{7}$$

Mapping. We map $\rho \in [0, 1]$ to a partial-count $M(\rho)$ and render a drone with M partials using additive synthesis (supported by AMY [26]). This establishes the interaction logic: *arrange nodes → infer topology → change a timbral regime.* Deploying and evaluating this mapping on hardware is ongoing work.

9 Discussion: Topology as a Compositional Medium

Sound Swarm contributes a compositional model in which relational structure, local behavior, and environmental coupling become primary materials for interactive music. The system extends prior swarm-music and Human–Swarm IMS work by treating spatial arrangement not as staging, but as a condition-setting practice through which sonic organization can be explored and shaped [2–4, 17, 18]. It also aligns with process-oriented and behavioral views of digital music systems, where composers specify conditions and interaction logics rather than fully determining event-level outcomes [5, 7]. For *Sound Swarm*, space is therefore not merely an output stage: the acoustic field stores and transforms traces through masking, reflections, and reverberation, which participate in subsequent behavior.

9.1 Topology quality as musical utility

The Stage (a) evaluation clarifies what "good topology" means for this instrument. Absolute localization is not the primary objective; the Ghost Map is used through neighborhood queries, density estimates, and cluster structure that parameterize local coupling. In the $N = 10$ ablation study, the combined variant achieves 10.03 ± 1.78 cm distance RMSE and 77.9% ± 8.8% kNN accuracy ($k = 2$) with an embedded-feasible model size (193 parameters, 0.75 KB). These results suggest that a minimal relational model can recover stable enough neighborhoods to support topology-conditioned behaviors under microcontroller constraints.

This utility is compositional as well as computational. The Ghost Map gives the composer an intelligible way to reason about adjacency, density, clustering, and exposure within the swarm. It invites experimentation with the acoustic field through rearrangement, listening, and iteration, shifting attention from hidden network state or event-level schedules toward conditions of coupling. Topology therefore becomes a creative medium for sonic exploration rather than a backstage technical layer [1, 13, 17, 18].

Positional RMSE remains variable under MDS, which we interpret as a design constraint: behaviors should be robust to global geometric distortion and rely primarily on local relational structure. In practice, topology-conditioned mappings should privilege *relative* features—nearest-neighbor sets, local densities, and cluster relations—over metrically precise coordinates. The map is valuable insofar as it supports interaction, not because it reconstructs exact geometry.

9.2 Ghost Maps as an interaction surface

By embedding calibration-derived pairwise sensing into a 2D Ghost Map, the swarm becomes legible and steerable as a population. This supports condition-setting: the composer or performer can inspect inferred relations, observe the effects of rearrangement, and intervene by tuning constraints rather than issuing event-level commands. *swarmtool* operationalizes this shift by exposing topology and agent state as a manipulable interface layer.

For interactive music practice, this frames engagement with the swarm neither as centralized sequencing nor as opaque emergence, but as the tuning of a distributed instrument whose behavior unfolds across population state, acoustic coupling, and time. The Ghost Map is more than a visualization of hidden data: it becomes a score-like representation of evolving relational possibility. It helps the composer reason about the swarm's current behavior, likely tendencies, and possible response to changed configuration. Historical states also matter, since prior inferred relations and configurations can inform future intervention. The resulting practice is close to what we describe as *spatial lutherie*: shaping the instrument by shaping the conditions under which it self-organizes [5, 16–18, 20].

9.3 Limitations

The main limitation is straightforward: the topology benchmarks reported here are synthetic. The strongest present claims are therefore conceptual and architectural: topology can function as a musically meaningful relational representation, bio-inspired local policies can be organized through a shared acoustic field, and *swarmtool* can make such systems legible as populations.

The framework is also informed by real prototyping. We have built and exercised v1 colonies at small scale ($N = 3$, $N = 6$, and $N = 10$), and the core firmware stack—including role election, telemetry, and pairwise calibration/probe exchange—is operational. The next step is to benchmark the same interaction loop on measured physical layouts and connect topology-conditioned behavior more directly to listener-facing outcomes. This will test how well the framework carries from synthetic verification into situated musical practice.

10 Conclusion

Sound Swarm presents an embodied, environmentally coupled mesh of synthesizer agents for composition via condition-setting rather than event specification. We describe a three-tier Perception–Behavior–Expression architecture, a calibration-driven relational perception loop (arrangement $\rightarrow$ tensor $\rightarrow$ Ghost Map), and an observability/control layer through *swarmtool*. Grounded in biological coordination, the Behavior layer operationalizes acoustic stigmergy, frog-inspired masking-avoidance timing, and lightweight role differentiation.

A simulation-first evaluation develops the topology inference pipeline from Classical MDS feasibility, through a calibrated baseline and lightweight learned distance regressor, to an $N = 10$ layout-split ablation. In the $N = 10$ setting, the combined RF+acoustic variant achieves 10.03 ± 1.78 cm distance RMSE and $77.9\% \pm 8.8\%$ neighborhood accuracy ($k = 2$) using a 193-parameter model (0.75 KB), suggesting that neighborhood-stable, topology-conditioned interaction is feasible under the synthetic measurement model while remaining compatible with embedded constraints.

Together, these contributions position *Sound Swarm* as a platform for exploring musical form as an emergent property of agent heuristics, spatial arrangement, and environmental acoustics. Next steps are Stage (b)/(c) tabletop and real-room benchmarking, sound examples, and expanded perception and behavior across diverse acoustic environments.

Acknowledgments

We thank Brian Whitman and the AMY open-source community for maintaining the AMY embedded synthesis engine used in this work [26].

11 Ethics Statement

This work follows the NIME Principles & Code of Practice on Ethical Research. The reported results did not involve a formal human-participant study. Nodes process only low-level, non-identifiable acoustic features locally; no personally identifiable audio, video, or tracking data were collected.

Biological references (e.g., honeybees, ants, and tree frogs) are used only as design heuristics for coordination and interaction; no biological experimentation or animal interaction was undertaken.

To support access and socio-economic fairness, the project uses relatively lightweight embedded hardware and is intended for open-source software release. We acknowledge that the current prototype has not yet been systematically evaluated for accessibility across diverse performer needs, and that embodied hardware may still limit replication and use.

Following NIME guidance on AI use, agentic coding tools provided limited support during software development, including notebook workflows and firmware prototyping. All resulting code and technical decisions were reviewed, validated, and integrated by the authors. Environmental impacts of hardware prototyping and electronic waste remain future work.

References

[1] Ikkyu Aihara, Daichi Kominami, Yasuharu Hirano, and Masayuki Murata. 2019. Mathematical Modelling and Application of Frog Choruses as an Autonomous Distributed Communication System. *Royal Society Open Science* 6, 1 (2019), 181117. https://doi.org/10.1098/rsos.181117

[2] Daniel Bisig, Martin Neukom, and John Flury. 2009. Interactive Swarm Orchestra: An Artificial Life Approach to Computer Music. In *Proceedings of the International Computer Music Conference (ICMC)*. ICMA, Montreal, Canada, 13 pages.

[3] Tim Blackwell. 2007. Swarming and Music. In *Evolutionary Computer Music*, Eduardo R. Miranda and James A. Biles (Eds.). Springer, London, 194–217. https://doi.org/10.1007/978-1-84628-600-1_9

[4] Tim Blackwell and Michael Young. 2004. Self-organised music. *Organised Sound* 9, 2 (2004), 123–136. https://doi.org/10.1017/S1355771804000214

[5] Oliver Bown, Alice Eldridge, and Jon McCormack. 2009. Understanding Interaction in Contemporary Digital Music: From Instruments to Behavioural Objects. *Organised Sound* 14, 2 (2009), 188–196. https://doi.org/10.1017/S1355771809000296

[6] Joel Chadabe. 1984. Interactive Composing: An Overview. *Computer Music Journal* 8, 1 (1984), 22–27. https://doi.org/10.2307/3679894

[7] Agostino Di Scipio. 2003. "Sound is the interface": from interactive to ecosystemic signal processing. *Organised Sound* 8, 3 (2003), 269–277. https://doi.org/10.1017/S1355771803000244

[8] Brian Eno. 1996. Generative Music. Talk delivered at the Imagination Conference, San Francisco; transcript published by In Motion Magazine. https://www.inmotionmagazine.com/eno1.html

[9] Sam Ferguson, Anthony Rowe, Oliver Bown, Liam Birtles, and Chris Bennewith. 2017. Sound Design for a System of 1000 Distributed Independent Audio-Visual Devices. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. NIME, Copenhagen, Denmark, 6 pages. https://doi.org/10.5281/zenodo.1176236

[10] Martin Gardner. 1970. Mathematical Games – The fantastic combinations of John Conway's new solitaire game "life". *Scientific American* 223, 4 (1970), 120–123.

[11] Simon Garnier, Jacques Gautrais, and Guy Theraulaz. 2007. The biological principles of swarm intelligence. *Swarm Intelligence* 1, 1 (2007), 3–31. https://doi.org/10.1007/s11721-007-0004-y

[12] Pierre-Paul Grassé. 1959. La reconstruction du nid et les coordinations inter-individuelles chez Bellicositermes natalensis et Cubitermes sp. la théorie de la stigmergie. *Insectes Sociaux* 6, 1 (1959), 41–80. https://doi.org/10.1007/BF02223791

[13] Michael D Greenfield and A. Stanley Rand. 2000. Frogs have rules: Selective attention algorithms regulate chorusing in *Physalaemus pustulosus* (Leptodactylidae). *Ethology* 106, 4 (2000), 331–347.

[14] Angel Hernández, Christian Blum, and Ikkyu Aihara. 2011. Self-organized frog chorus as an example of distributed graph coloring. In *Computer Aided Systems Theory – EUROCAST 2011 (Lecture Notes in Computer Science, Vol. 6927)*. Springer, Berlin, Heidelberg, 165–172. https://doi.org/10.1007/978-3-642-23960-1_21

[15] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. 2018. Neural Relational Inference for Interacting Systems. In *Proc. 35th Int. Conf. on Machine Learning (ICML) (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, Stockholm, Sweden, 2688–2697.

[16] Michael Krzyżaniak. 2021. Musical robot swarms, timing, and equilibria. *Journal of New Music Research* 50, 3 (2021), 279–297. https://doi.org/10.1080/09298215.2021.1910313

[17] Pedro Lucas and Kyrre Glette. 2023. Human-Swarm Interactive Music Systems: Design, Algorithms, Technologies, Evaluation. In *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research (CMMR 2023)*. CMMR, Tokyo, Japan, 12 pages.

[18] Pedro Pablo Lucas, Alexander Szorkovszky, Stefano Fasciani, and Kyrre Glette. 2024. HS-ims: A Platform for Human-Swarm Interactive Music Systems. In *Proc. IEEE Int'l Conf. on Autonomic Computing and Self-Organizing Systems (ACSOS)*. IEEE, Aarhus, Denmark, 145–150. https://doi.org/10.1109/ACSOS61780.2024.00034

[19] Ryan A. McCarthy, You Zhang, Samuel A. Verburg, William F. Jenkins, and Peter Gerstoft. 2025. Machine Learning in Acoustics: A Review and Open-source Repository. *npj Acoustics* 1, 1 (2025), 18. https://doi.org/10.1038/s44384-025-00021-w arXiv:2507.04419.

[20] Soichiro Mihara and Kazuki Saita. 2019. Moids 2.2.0 — acoustic emergence structure. In *Sound Art: Sound as a Medium of Art*, Peter Weibel (Ed.). The MIT Press, Cambridge, MA, 340–341. Sound installation (2009/2012). Conceptual adviser: Hiroko Mugibayashi.

[21] Craig W. Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. *Computer Graphics* 21, 4 (1987), 25–34. https://doi.org/10.1145/37401.37406 ACM SIGGRAPH '87 Conference Proceedings.

[22] Alessandro Ricci, Andrea Omicini, Mirko Viroli, Luca Gardelli, and Enrico Oliva. 2007. Cognitive Stigmergy: Towards a Framework Based on Agents and Artifacts. In *Environments for Multi-Agent Systems III*. Lecture Notes in Computer Science, Vol. 4389. Springer, Berlin, Heidelberg, 124–140. https://doi.org/10.1007/978-3-540-71103-2_7

[23] R. Murray Schafer. 1994. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Destiny Books, Rochester, VT. Originally published as *The Tuning of the World*, New York: Knopf, 1977.

[24] Thomas D. Seeley. 2010. *Honeybee Democracy*. Princeton University Press, Princeton, NJ.

[25] Brian Whitman. 2020. Alles: A many speaker distributed music synthesizer using UDP multicast over WiFi. Open-source software. https://github.com/shorepine/alles

[26] Brian Whitman. 2023. AMY – A high-performance fixed-point music synthesizer library for microcontrollers. Open-source software. https://github.com/shorepine/amy