Investigation of the Cellular Au-Tonnetz as an intermedial composition tool

Abstract

Electroacoustic audiovisual composition often aspires to parity between sound and moving image, yet compositional workflows and perceptual salience frequently reintroduce hierarchy, positioning one medium as structural and the other as illustrative. Responding to calls to re-conceptualise media relationships through intermediality, this article examines intermedial interference: constructive and destructive interactions between media features, as an evaluable property of generative audiovisual instruments. We study a Tonnetz cellular automata (CA) system in which sound and LED-panel activity are parallel projections of a single evolving state-space. Participants shaped the system by adjusting a compact parameter set governing population bounds, neighbourhood extent, neighbour-rule thresholds, update rate, activation lifetimes, recurrence (looping), and harmonic admissibility (scale). Using a within-subject design, each participant completed three short compositional tasks under controlled modality-access conditions (visual-only, audio-only, audiovisual). In all conditions, both audio and video were recorded and end-state parameter presets saved. Participants reported satisfaction, intention, steerability, intermedial coherence, balance, autonomy, and interference, followed by immediate replay-based reflection with both modalities present. We outline an analysis framework linking compositional experience to captured audiovisual artefacts and propose an interference taxonomy grounded in parameter regimes and observed strategies (e.g., density shaping, looping, and scale manipulation). A blinded external rating study of the resulting clips is underway to triangulate participant experience with artefact-level judgements. The study offers a practical evaluation method for practice-based intermedial research and a state-space projection framing of audiovisual composition as parallel media projections of a shared generative substrate.

Keywords: intermediality; audiovisual composition; electroacoustic music; generative systems; cellular automata; Tonnetz; media equality; intermedial interference; co-creative agency; practice-based

1 Introduction

Electroacoustic audiovisual composition is often discussed through models that implicitly privilege one medium over another: sound interpreted as primary structure with image as illustration, or image interpreted as primary narrative with sound as enhancement. Even where practitioners aim for parity, the practical realities of compositional workflow, interface design and perceptual salience tend to reintroduce hierarchy. This thematic issue calls for approaches that treat sound and moving image as co-equal media within an intermedial space, and for critical accounts of how meaning arises when the media are fused rather than juxtaposed. This article contributes; a practice-based, empirical study of intermedial composition using a Tonnetz cellular automata (CA) system in which sound and light are generated from a single evolving state-space. The work starts from a simple proposition: if sound and image are projections of the same latent “score”, then intermediality can be addressed not only as an aesthetic aspiration, but as a manipulable design problem. In such a system, relationships between media are not primarily established through post-hoc synchronisation or representational alignment, but through shared generative causality: changes to the underlying state propagate to both modalities. This framing has two implications for intermedial practice. First, it invites a conception of composition as shaping the dynamics of an underlying state-space rather than authoring separate sonic and visual layers. Second, it provides a concrete basis for analysing intermedial interference: constructive and destructive interactions can be observed when the two projections amplify, compete with, or contradict one another in perception.

1.1. Intermedial interference as a design and evaluation problem In the context of generative audiovisual instruments

Intermedial interference can be understood as the perceptual and interpretive consequences of interactions between salient features of the two media: temporal alignment (or misalignment), structural correspondence (or divergence), salience dominance, and the distribution of perceived agency between human and system. In a state-space projection model, interference is not an incidental by-product; it is an emergent property of coupling choices, control affordances and constraint regimes. For example, increasing the update rate or density of events may raise perceived vitality while simultaneously producing attentional overload, shifting the balance of salience between modalities. Similarly, introducing recurrence (looping) may increase structural legibility while reducing the sense of liveness or co-authorship. These tensions are compositional in nature, but they are also evaluable: they manifest in what participants do, what they report experiencing, and what independent listeners/viewers perceive in the resulting artefacts. Despite the prevalence of generative systems in electroacoustic

and audiovisual practices, evaluation is frequently limited to descriptive accounts or anecdotal responses, making it difficult to compare design choices or to articulate how a system supports co-equal media relationships. There remains a need for evaluation approaches that can (i) retain practice-based validity, (ii) produce analysable evidence within realistic artistic development timelines, and (iii) connect participant experience to perceivable properties of the resulting audiovisual artefacts.

1.2. System overview: The Cellular Au-Tonnetz as an intermedial score.

The system studied here organises pitch classes on a Tonnetz-derived spatial lattice and drives their activation through asynchronous CA dynamics. The same evolving CA state is rendered simultaneously as (i) audio output and (ii) LED-panel activity, with fixed mappings between pitch-class admissibility and colour. Composition occurs through real-time adjustment of a small set of parameters controlling population bounds, neighbour-rule thresholds, temporal rate, activation lifetimes, neighbourhood extent (local versus extended), recurrence (looping), and harmonic admissibility (scale). Rather than treating sound and light as separate channels to be aligned, the instrument treats them as co-equal surfaces of a single evolving substrate: when the underlying state changes, both modalities change. This design intentionally blurs boundaries between instrument and environment. On one hand, it affords local interventions and steerability through parameter changes; on the other, it produces autonomous evolution that can surprise and resist direct control. The central evaluation question is therefore not whether the system is “usable” in a conventional sense, but how its coupling and constraints support (or undermine) intermedial equality, meaningful fusion, and negotiated authorship.

1.3. Research questions

The study addresses the following questions:

RQ1 Intermedial equality: To what extent do participants experience sound and light as co-equal contributors to meaning when composing with the system?

RQ2 Intermedial interference: What forms of constructive and destructive intermedial interference arise in practice, and how do they relate to modality access and parameter shaping?

RQ3 Agency and causality: How do participants’ perceptions of authorship, predictability and steerability change when one modality is unavailable during composition?

1.4. Methodological approach

To operationalise “co-equal media” without reducing the work to preference testing, the study uses a within-subject design in which each participant completes three short compositional tasks under controlled modality-access conditions: visual-only, audio-only and audiovisual. In all cases, both audio and video are recorded, end-state parameter presets are saved for reproducibility, and participants complete brief post-block measures targeting coherence, novelty, agency, fusion/equality, and constructive versus destructive interference. After each block, participants replay the captured excerpt with both modalities enabled and provide a short reflection on expectation, fusion and competition. A baseline “system-alone” capture provides a control for the system’s intrinsic generativity. To relate felt experience to perceivable outcomes, the captured clips are subsequently rated by independent raters drawn from the author’s wider network using blind presentation and standardised scales (preference, coherence, novelty, fusion/equality, overload). Finally, an exploratory dyad variant tests distributed access: one participant composes with audio-only feedback while the other composes with visual-only feedback, negotiating parameter changes verbally to reach a joint outcome.

1.5. Contributions

This article offers three contributions to intermedial electroacoustic audiovisual practice:

1. A state-space projection framing of intermedial composition in which sound and light are treated as parallel projections of a single generative substrate (an ‘intermedial score’), supporting analysis without default media hierarchy.
2. An evaluative method suitable for practice-based research under time constraints, combining modality-access manipulation, replay-based reflection, baseline controls and independent rater assessment to connect compositional experience to perceivable artefacts.
3. An interference-focused analytic account (taxonomy and exemplars) describing how constructive and destructive intermedial interference manifests in this system, and how parameter regimes and modality access relate to perceived fusion, agency and overload.

1.6. Article structure

Section 2 situates the work within intermedial audiovisual composition and generative electroacoustic practices, clarifying how interference and equality are understood in this study. Section 3 details the system and study method, including parameter set, capture

protocol and rating procedures. Section 4 reports quantitative and qualitative findings, including within-subject condition effects, baseline comparisons and rater triangulation, alongside an interference taxonomy supported by audiovisual exemplars. Section 5 discusses implications for intermedial instrument design, including practical strategies for balancing autonomy, constraint and salience to support co-equal media relationships. The audiovisual instrument evaluated here is the TZ5 / Cellular Au-Tonnetz system, previously described in detail as a unified audio-visual MIDI generator that couples Tonnetz pitch-space mapping with cellular automata and a web-configured embedded control architecture. (Didiot-Cook, 2025). That prior publication focuses on system design and demonstration rather than formal evaluation. The present paper therefore contributes an issue-driven evaluation aligned with Organised Sound's call for intermediality: specifically, how "media equality" and intermedial interference manifest when participants compose under modality-restricted conditions (visual-only / audio-only / audiovisual) and when outputs are subsequently assessed by independent raters.

2 Related work

2.1 Intermediality and "media equality" as a relational claim

Intermediality is often invoked to describe work that occurs "between" media, but its analytic value depends on specifying which intermedial relation is at stake. Rajewsky treats intermediality as a research lens and proposes a typology (medial transposition, media combination, intermedial reference) that helps avoid collapsing distinct practices into a single "multimedia" category (Rajewsky 2005). In this paper, the focus is media combination; sound and time-based light output co-present, and, more specifically, whether participants and observers experience the combined outcome as co-constitutive rather than one medium functioning as explanatory support for the other. The Organised Sound call sharpened this as a "media equality" problem: how meaning is produced when media are fused rather than merely juxtaposed (Organised Sound 2026).

Elleström's modality model provides a useful scaffold for discussing equality without treating media as undifferentiated streams. By distinguishing material, sensorial, spatiotemporal, and semiotic modalities, the model makes it possible to describe where equality succeeds or fails: parity may hold in causal origin (both derive from the same substrate) but collapse in sensorial salience (one dominates attention) or semiotic authority (one is treated as the "real" carrier of form) (Elleström 2010). The present study adopts this precision: "equality" is treated as negotiated and contingent, not assumed.

2.2 Audiovisual relations: integration, framing, and interpretive hierarchy

Audiovisual theory in film and sound studies provides durable concepts for how sound and image are perceived as an integrated whole. Chion's account of audio-vision emphasises that audiovisual perception is not additive; sound and image reciprocally frame one another's meaning (Chion 1994). This is directly relevant to a modality-withholding design: composing with one modality absent and reinstating it at replay makes the construction of audio-vision observable, including how expectation, surprise, and retrospective binding contribute to perceived unity.

Intermedial frameworks also emphasise analysing relations between media, in so far as how they align, diverge, reinforce, or contradict, ather than presuming stable hierarchies such as accompaniment/illustration (Rajewsky 2005; Elleström 2010). These accounts legitimise treating audience interpretation as primary evidence for intermedial success, and they align with the paper's focus on interference effects that emerge as relational phenomena rather than technical synchronisation achievements.

2.3 Intermedial interference as an evaluable property of coupled instruments

The Organised Sound call foregrounds "intermedial interference" as a prompt for reconceptualising media relationships, including constructive and destructive interactions between media features (Organised Sound 2026). Chiaramonte develops the term within electroacoustic audiovisual composition, treating the intermedial interaction itself as the compositional site and noting the difficulty of measurability when interference is phenomenological and context-dependent (Chiaramonte 2024).

The present study extends this line by focusing on a coupled generative instrument in which sound and light are parallel media projections of a single evolving state-space. This coupling makes interference design-sensitive: parameter regimes, constraint conditions, and control affordances can shift interference from productive fusion (reinforcement, heightened legibility, surprise-as-rebinding) to destructive conflict (contradictory cues, masking, reduced causal confidence). The within-subject modality-access manipulation is used to render these shifts tractable in practice-based evaluation.

2.4 Multisensory binding, crossmodal correspondences, and “cue” mechanisms

Work in multisensory perception suggests that binding depends on which cues are treated as reliable evidence for a shared causal story. Crossmodal correspondences offer a complementary design lens: Spence’s tutorial review synthesises evidence that people reliably associate features across modalities (e.g., pitch–brightness), making correspondences a resource for designing legible mappings rather than an anecdotal curiosity (Spence 2011). For the present instrument, correspondences help explain why some participants may use colour similarity as a proxy for harmonic planning, and why constraint conditions can re-weight which cues dominate attention and inference.

2.5 Agency, co-creativity, and evaluation traditions in DMIs and generative systems

Generative and autonomous instruments often trade direct predictability for novelty and emergence, prompting an agency negotiation in which performers steer rather than command. Evaluation traditions in Digital Musical Instruments (DMIs) provide precedents for studying this without reducing outcomes to usability metrics. O’Modhrain proposes an evaluation framework for DMIs that supports structured reflection on experience and capability while remaining sensitive to musical context (O’Modhrain 2011). Orio and Wanderley similarly discuss evaluation of input devices for musical expression, offering method rationales that bridge performer experience and system design iteration (Orio and Wanderley 2002).

2.6 Tonnetz and cellular automata as compositional substrates

The Tonnetz has a substantial theoretical lineage within neo-Riemannian and transformational theory, providing a well-motivated harmonic space in which neighbourhood relations correspond to parsimonious voice-leading transformations. Cohn’s account of neo-Riemannian operations and Tonnetz representations provides a canonical grounding for treating Tonnetz geometry as more than a “clever layout” (Cohn 1997). Douthett and Steinbach’s work on parsimonious graphs further reinforces the status of these geometries as analytical and compositional structures rather than purely visualisations (Douthett and Steinbach 1998).

Besada et al. (2024) highlight cognitive/HCI challenges in ‘Tonnetz-at-first-sight’ interaction, where visually salient geometric regularities can bias harmonic expectations. This study treats such bias not only as a usability risk but as a potential

resource: a site where ‘misleading’ structure can be reframed as productive intermedial interference, supporting exploratory interpretation in a coupled audio–light system

Cellular automata (CA) also have a long-standing lineage as musical generative formalisms, valued for producing emergent structure from simple local rules. Miranda’s chapter on cellular automata music provides a clear anchor for positioning CA as a legitimate compositional engine spanning sound generation and higher-level form (Miranda 2007). In the present system, the distinctive move is to treat CA state as an intermedial score: the same evolving substrate is rendered simultaneously as sound and as time-based light activity, enabling interference to be observed as an emergent property of coupled projections under constraint.

3 Method

3.1 Participants and ethics

Recruitment and sampling. Participants were recruited via the author's professional and personal networks using convenience sampling. Recruitment aimed to include a range of musical experience and varying familiarity with generative systems.

Solo sample. Nine participants completed the solo protocol (N = 9), each completing all three modality-access blocks (27/27 blocks total). Participants self-reported age range, musical experience, music-theory familiarity, familiarity with generative systems, Tonnetz familiarity, and perceptual notes (colour-vision deficiency, sensitivity to flashing lights, and optional comments). One participant reported red–green colour deficiency; one participant used speakers rather than headphones during the audio-only block.

Exploratory dyad. In addition to the solo protocol, one exploratory dyad trial was conducted in which the same pair completed two dyad sessions with role-swap (audio-role vs visual-role), yielding two dyad captures. Dyad data are treated as exploratory and are reported descriptively.

Consent and recording. All participants provided informed consent for audio/video capture and for anonymised excerpts to be used as research stimuli (including for an external rating task described below as ongoing).

3.2 System and apparatus

3.2.1 Instrument under study

The study evaluates the Cellular Au-Tonnetz (TZ5) audiovisual instrument, a coupled generative system in which sound and LED-panel activity are produced as parallel media projections of a single evolving Tonnetz-cellular automata (CA) state.

Figure 1 shows the study system diagram, comprising the TZ5 and its constituent parts, as well as the supporting infrastructure and instrumentation required to conduct the study.

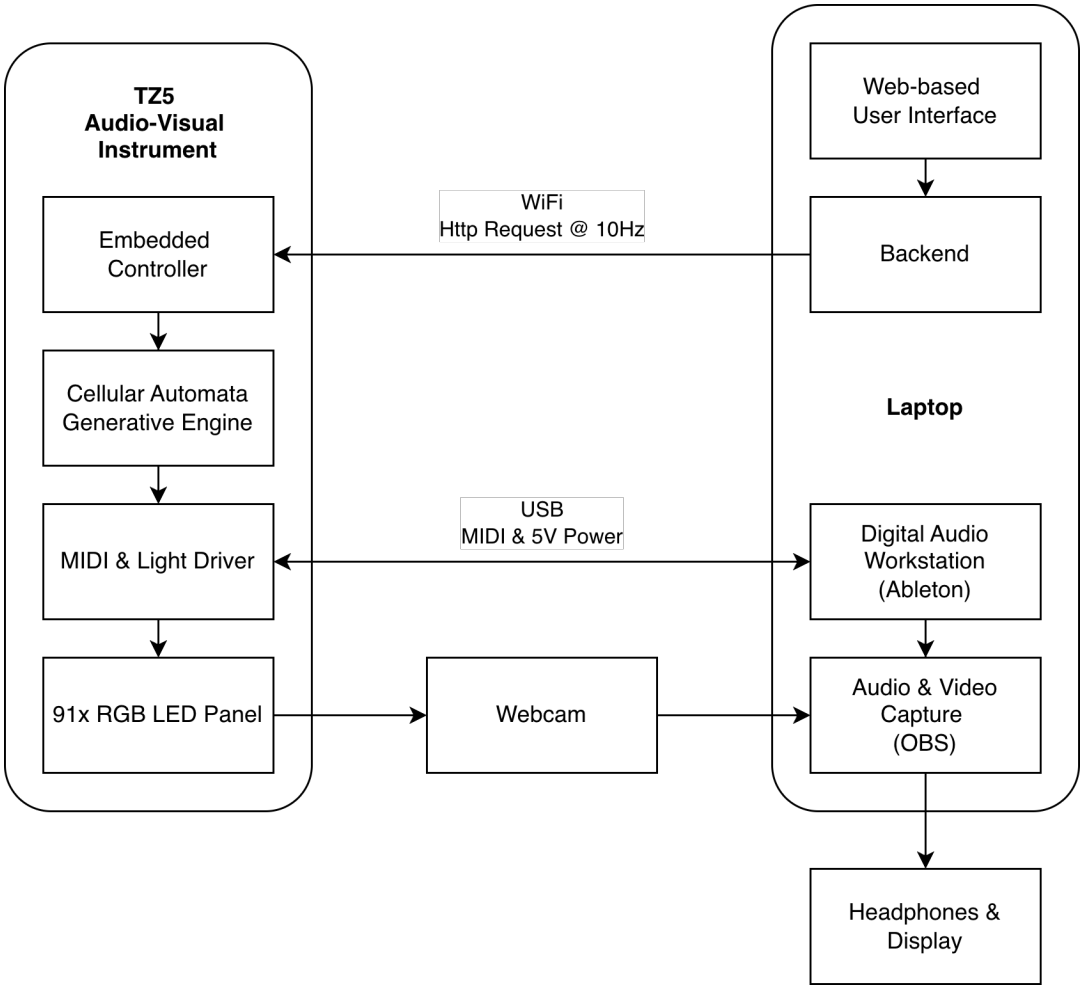

Figure 1. System block diagram (UI → ESP32-S3 → LED panel + MIDI/audio chain).

The instrument comprises four functional components:

1. Web control interface and configuration store. A browser-based UI provides sliders and toggles for the study parameter set. Parameter values are stored in an SQL-backed configuration store and served to the device as a JSON configuration payload over HTTP.

2. Embedded controller. An ESP32-S3 microcontroller retrieves the latest configuration, updates the CA state, and drives both the audio-control stream and LED output in real time.

3. LED panel. A 91-pixel hexagonal LED array displays the evolving CA state using a fixed pitch-class–to-colour mapping.
4. Audio rendering chain. The embedded controller outputs musical events as MIDI, rendered using a fixed audio synthesis configuration (Ableton Live with a single sine tone MIDI instrument) held constant across sessions.

Figure 2 shows an exploded view of the TZ5 LED panel and Embedded controller. The controller connects over USB C to the laptop for power and MIDI transmission.

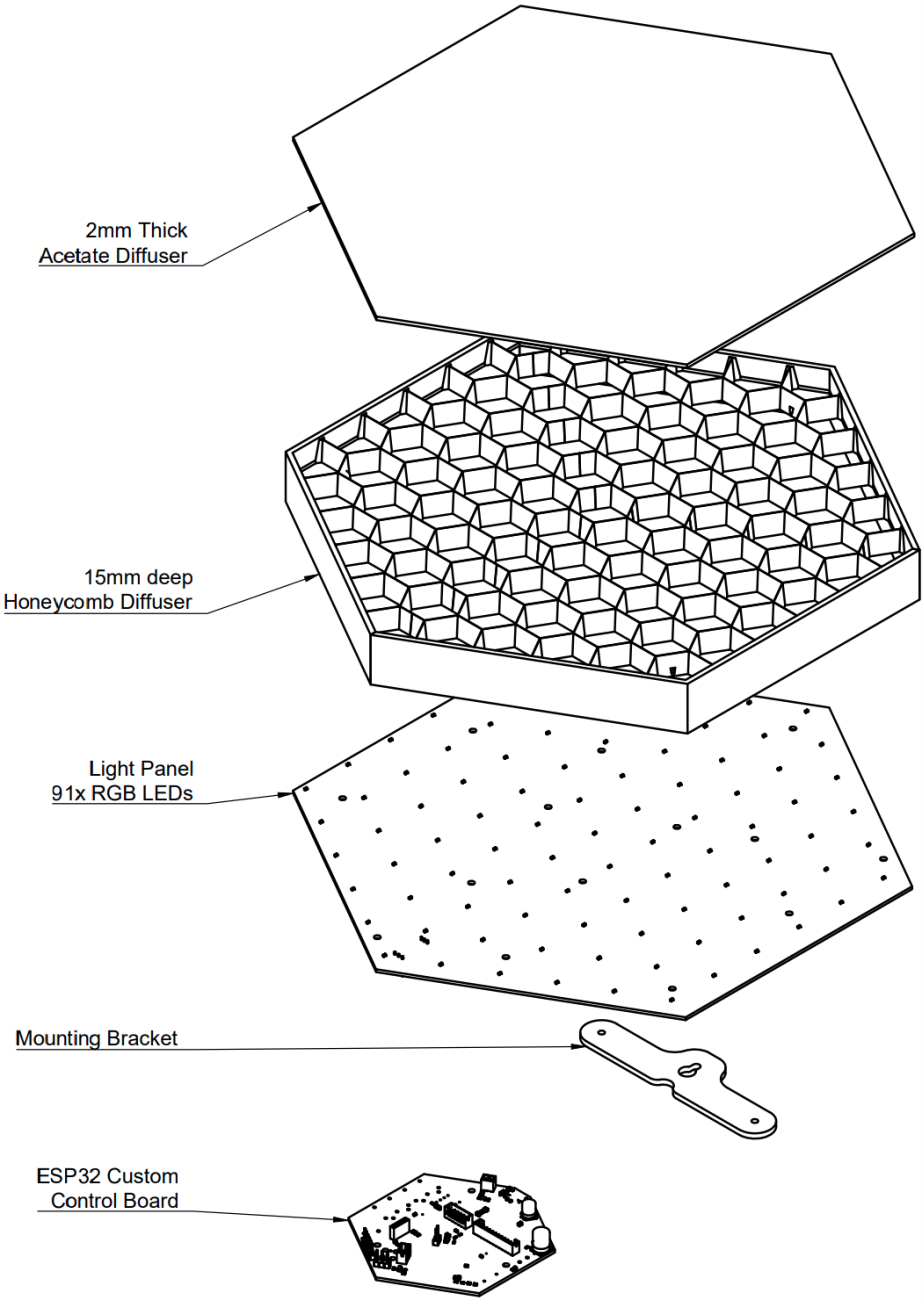

Figure 2. Exploded view of TZ5 audiovisual instrument. Light from the 91x Red, Green & Blue (RGB) LEDs is guided by a hexagonal lattice and diffused by a 2mm sheet of acetate.

3.2.2 Coupled generation and mappings

The system maintains a Tonnetz-mapped lattice of pitch classes. At each update, local CA rules and global population constraints determine which cells become active. When a cell activates, the system emits (i) a musical event (e.g., MIDI note) and (ii) a corresponding LED update. In this sense, sound and light are coupled by shared causal origin (state update), not by post-hoc synchronisation.

Pitch classes are mapped to hue (HSV-based mapping), such that scale constraints simultaneously affect the admissible pitch set and the resulting colour palette. Colour mapping and synthesis settings were held constant across participants to keep perceptual comparisons focused on state dynamics and parameter steering rather than renderer variability.

Figure 3 shows the tonnetz pitch class and hue mapping. Hue increments around the circle of fifths, and MIDI number increments from the lowest note C;36 to the highest A#;106.

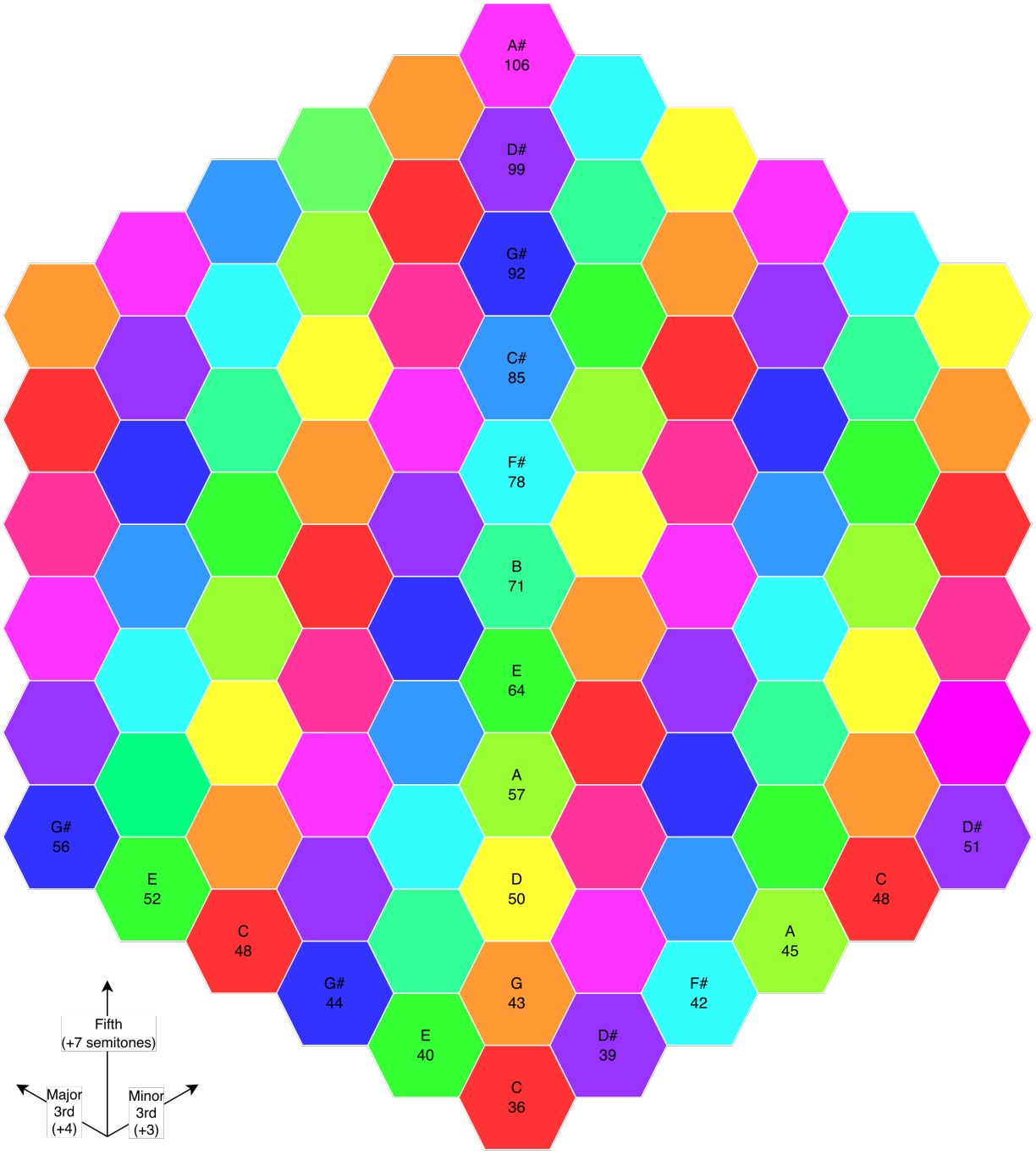

Figure 3 Tonnetz pitch-class map used for hue mapping. MIDI numbers are given under each pitch class, and computed from the root of C;36 (bottom, centre), incrementing by +3 semitones moving right, and +4 moving left.

3.2.3 LED panel configuration (study setup)

The study used a 91-LED hex panel with fixed diffusion and fixed participant viewing geometry, held constant across all sessions.

3.2.4 Participant-facing control parameters

Participants controlled the instrument exclusively via the web UI. The study parameter set targeted population/density, temporal dynamics, neighbourhood rule constraints, recurrence, and harmonic admissibility:

- Population and density
- Max Population (Max Notes): upper bound on concurrent active cells.
- Min Population (Min Notes): lower bound; when activity drops below this threshold the system reseeds activity to maintain baseline density.
- Temporal dynamics
- Rate: CA update rate (ms or equivalent).
- Life Length: activation lifetime parameter (firmware-defined units).
- Neighbourhood and rule constraints
- Neighbourhood extent: Local (6 neighbours) vs Extended (18 neighbours).
- Min Neighbours / Max Neighbours: activation thresholds and crowding constraint.
- Recurrence and memory
- Loop On/Off: toggles between stochastic selection and replay of a stored coordinate-selection sequence.
- Loop Length (Loop Steps): number of update steps stored and replayed; effective loop duration depends on Rate × Loop Length.
- Harmonic admissibility
- Scale: constrains which pitch classes can activate; because pitch classes map to hue, scale selection also shapes the visible palette.

Automated chord progression / automatic scale sequencing features were disabled for this study, to attribute formal change primarily to participant steering and the coupled CA dynamics.

3.3 Study design

3.3.1 Overview (solo protocol)

The solo study used a within-subject design with three short composition blocks per participant, each manipulating modality access during composition:

A: Visual-only (panel visible; audio masked)

B: Audio-only (audio via headphones/speakers; panel occluded)

C: Audiovisual (both modalities available)

In all conditions, both audio and video were captured, and end-state parameter presets were saved to support traceability and repeatability.

Figure 4 shows the studio set up, detailing the instrument and the supporting infrastructure.

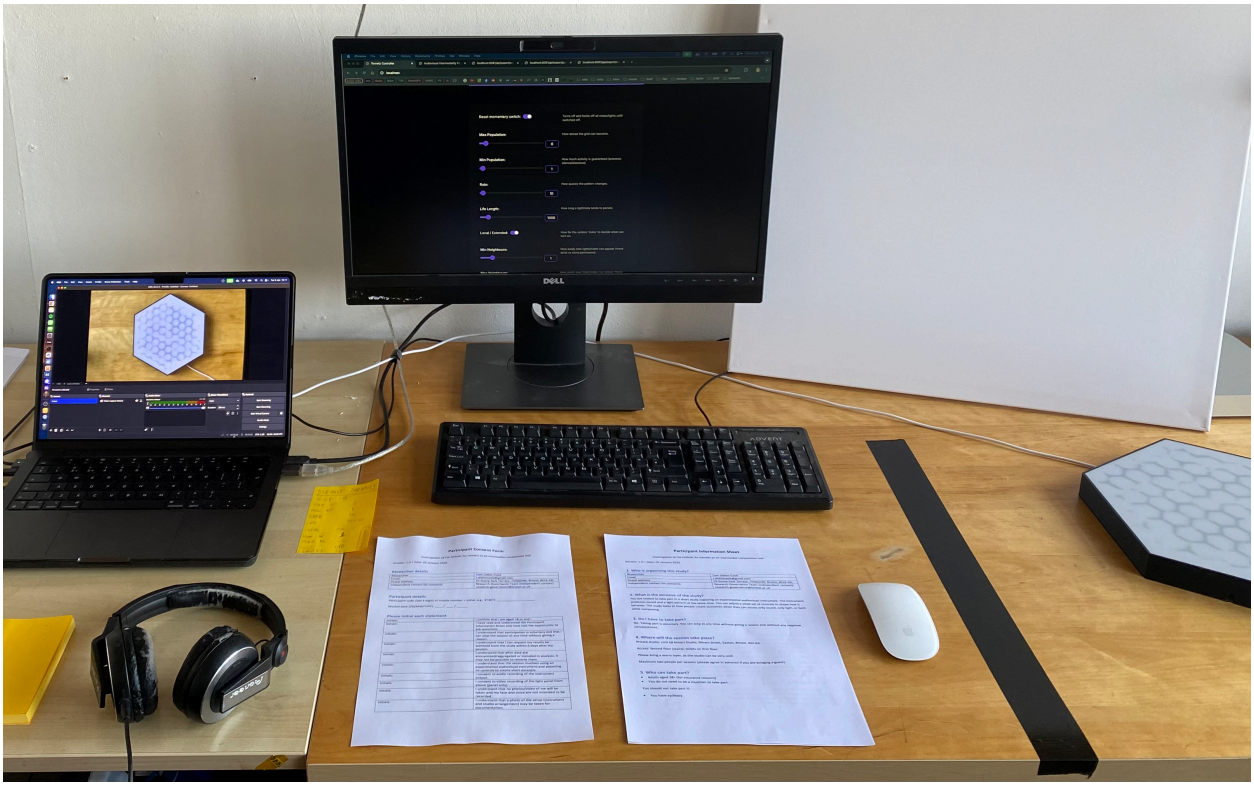

Figure 4 - Experiment setup (from left to right); Laptop hosting interface backend and recording software, Headphones, secondary display for survey and web control interface, consent form and participant information sheet, canvas for obscuring instrument, TZ5 instrument. Out of view in the top right is the webcam looking down on the TZ5.

3.3.2 Order control

Block order was counterbalanced across participants using a rotating schedule to reduce systematic order effects. Participant ID and block order were recorded in session metadata.

3.3.3 Baseline control (system-alone)

To characterise intrinsic generativity absent human steering, the system was also captured running from the common starting preset **S0** with no parameter changes for the same duration as a participant block. This produced a baseline audiovisual artefact for qualitative comparison and for optional inclusion in the external rating stimulus set.

Figure 5 shows the web control interface with parameters set to starting preset **S0**.

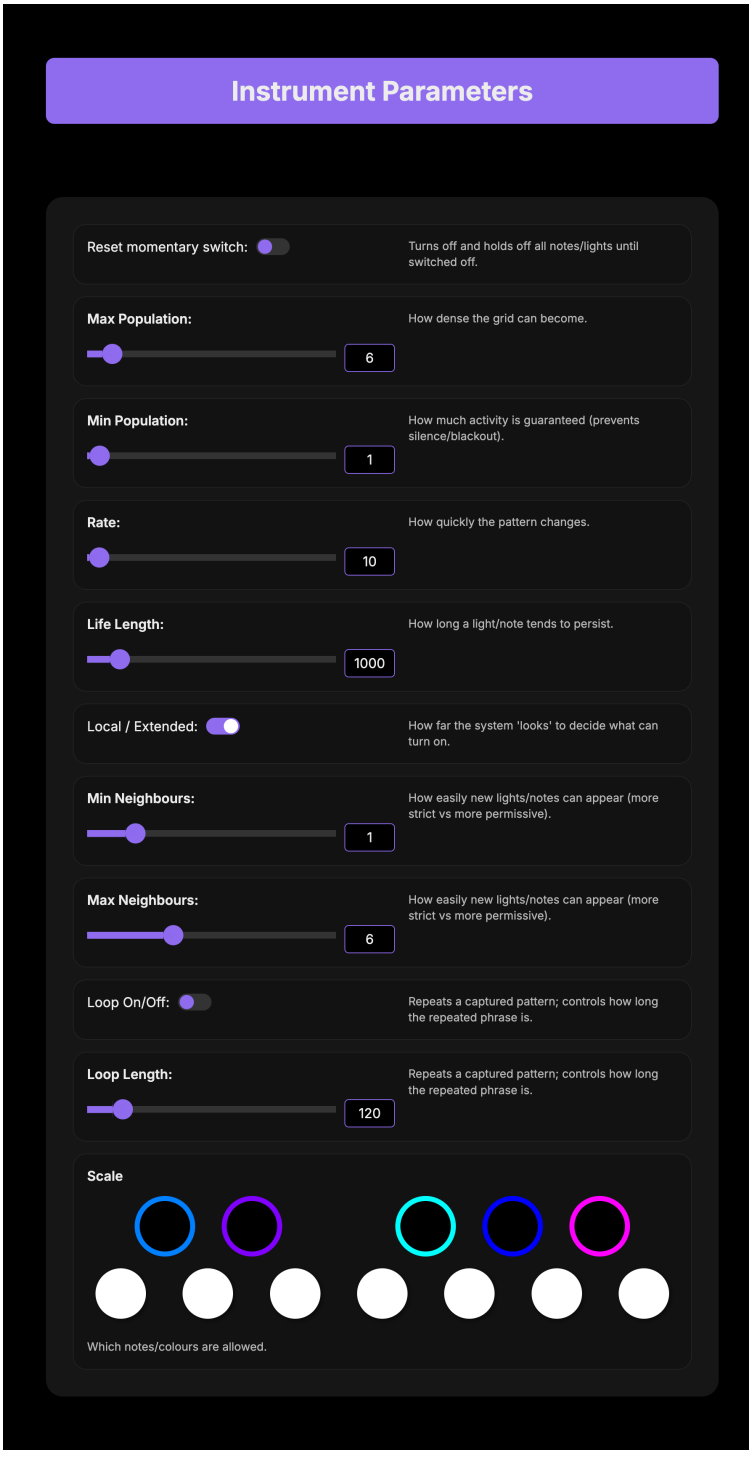

Figure 5 - Control interface screenshot set to common starting preset (**S0**)

3.3.4 Exploratory dyad protocol (asymmetric modality access)

A single pair completed two dyad sessions with role-swap. In each dyad session, one participant had auditory access while the other had visual access, and the pair negotiated parameter changes verbally:

1. Audio-role: auditory access (headphones); no view of the panel.

2. Visual-role: panel view; no access to audio.

The dyad produced a final captured excerpt from **S0** after collaborative steering. Dyad findings are treated as exploratory and used as a “teaser” for future systematic study.

3.4 Procedure

Figure 6 shows the flow diagram for the study procedure. From recruitment to consent, completion of the 3 composition, reveal & questionnaire blocks (audio only, visual only and with access to both modalities) to final reflections. More detail on each step is available in the following sub sections.

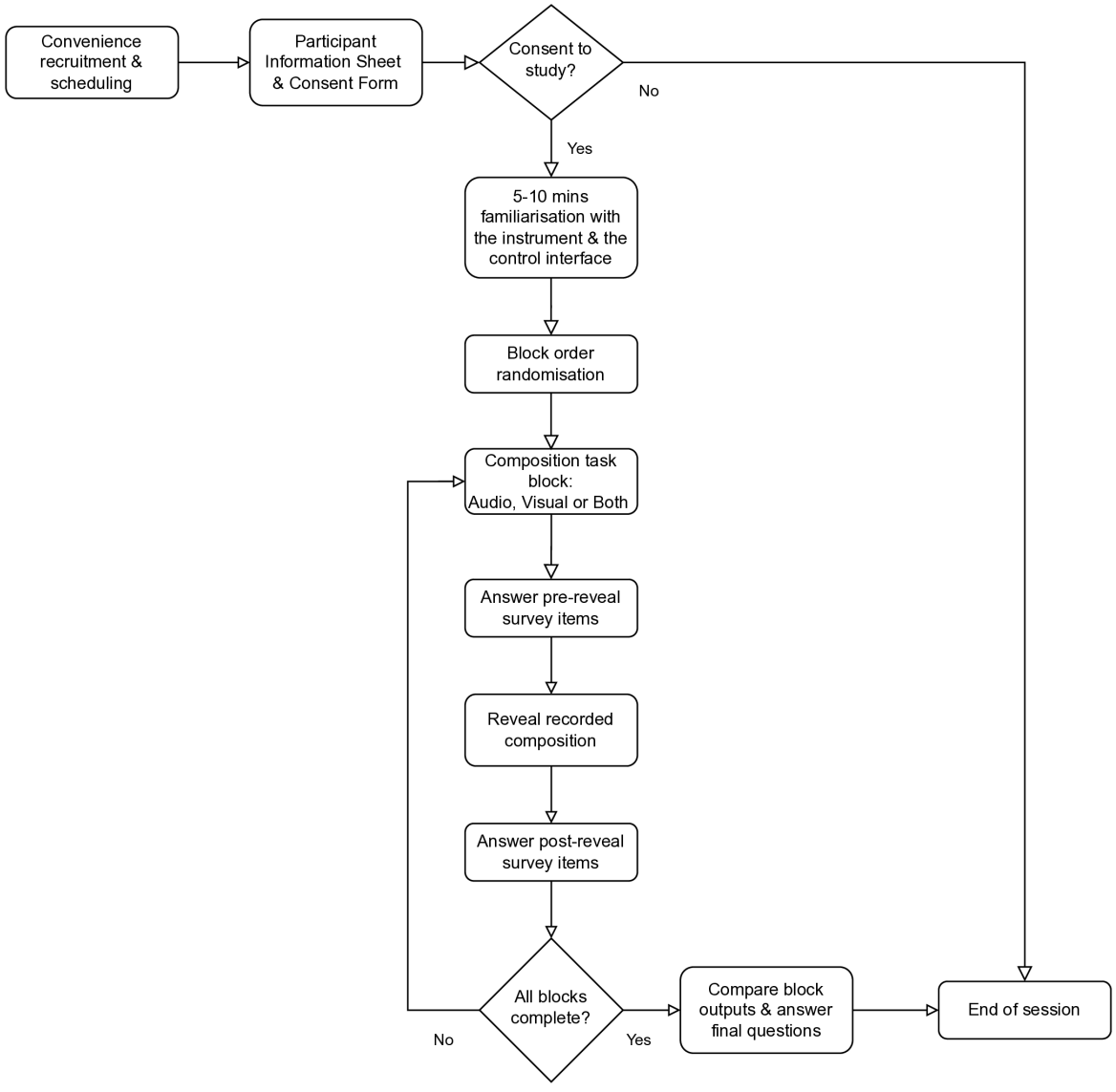

Figure 6. Study procedure flow diagram. All 9 participants completed the full procedure.

3.4.1 Orientation and familiarisation

Participants received a short briefing describing the instrument as a coupled audiovisual generator and outlining the three blocks. A familiarisation period (5-10 minutes) with full audiovisual access introduced the core parameters (density bounds, rate, neighbourhood extent, looping, and scale constraints).

3.4.2 Common starting point and captures

Each block began from a standardised preset **S0** to support within-subject comparison. Participants explored the system under the assigned modality constraint for ~5 minutes, then recorded a final excerpt of [60–120] seconds. Parameter changes were permitted during recording. The end-state preset for each block was saved.

Block prompts were:

A (Visual-only): “Create an interesting pattern/motion on the LED panel.”

B (Audio-only): “Create an interesting melody/sonic texture.”

C (Audiovisual): “Create an interesting combined audiovisual experience.”

3.4.3 Reveal and per-block questionnaires

After each capture participants the excerpt was replayed with both modalities enabled (the “reveal”). Participants completed:

- Part A (pre-reveal): ratings and brief notes about the composition experience under constraint.
- Part B (post-reveal): ratings and reflections about the fused audiovisual result after replay.

This structure was designed to separate (i) compositional experience under constraint from (ii) retrospective binding and reinterpretation when both modalities are present.

3.4.4 Dyad trial procedure

For each dyad session: confirm asymmetric access, explore from S0 (~5–10 minutes), record a final excerpt ([60–120] seconds) with parameter changes permitted, save the end-state preset, and complete the dyad questionnaire independently.

3.5 Measures

3.5.1 Per-block questionnaires (solo)

Each block used a two-part questionnaire with 7-point Likert ratings (1 = strongly disagree, 7 = strongly agree), plus short free-text prompts.

Part A (pre-reveal) captured:

- satisfaction with outcome (A1),
- intention clarity (A2),
- steerability toward intention (A3),
- interface understandability (A4),
- useful surprise (A5),
- frustrating unpredictability (A6),
- confidence others would find the result interesting (A7), plus brief strategy notes.

Part B (post-reveal) captured:

- same-process judgement (B1),
- modality balance (B2),
- coherence/legibility (B3),
- constructive reinforcement (B4),
- destructive contradiction (B5),
- overload (B6),
- expectation match (B7),
- interpretation change (B8),
- causal story plausibility (B9),
- perceived system autonomy (B10),
- reliance on visual cues (B11),

- reliance on Tonnetz/music-theory cues (B12), plus free-text reflections on mismatch, fusion, and interference moments.
- Participants also nominated up to three “most influential” parameters used in that block.

3.5.2 End-of-session comparison (solo)

At session end, participants:

- ranked their three outputs,
- selected which block felt “most intermedial,”
- selected which block had the “biggest mismatch,”
- provided one change request to improve media equality / intermedial legibility.

3.5.3 Dyad questionnaire (exploratory)

Dyad participants completed a short questionnaire independently, rating coordination, communication, shared reference points, perceived audiovisual coherence/balance, and joint ownership, alongside brief notes on negotiation and disagreements.

3.5.4 External ratings (planned triangulation; ongoing)

A separate pool of independent raters was recruited to evaluate anonymised audiovisual clips in a blinded, randomised web presentation. The intent of this rating task is to triangulate performer self-report with observer judgements of the captured artefacts. At the time of writing, this rating process is ongoing; therefore, the present paper reports no inferential claims from rater data. Rating progress (number of raters and clip completions) is documented in an appendix.

3.6 Data handling and analysis plan

3.6.1 Captured artefacts per solo block

- end-state parameter preset (machine-readable export),
- video file for the captured excerpt,
- Part A and Part B questionnaire responses,
- end-of-session ranking and reflections.

- within-subject contrasts across A (visual-only), B (audio-only), and C (audiovisual),
- exploratory description of dyad behaviour and experience.

3.6.2 Qualitative analysis focus

- compositional strategies under constraint,
- reveal-driven reinterpretation and mismatch valence,
- interference mechanisms (e.g., temporal/density cue conflicts, masking),
- participant-led design requirements.

3.6.3 Quantitative analysis focus

- condition-level summaries of key Likert items,
- paired within-subject visualisation of focal measures (e.g., steerability, same-process judgement, overload),
- descriptive distributions of end-of-session choices (preference, “most intermedial,” “biggest mismatch”),
- descriptive distributions of parameter nominations by condition.

4 Results: composing and interpreting intermedial relations under modality constraint

This section reports within-subject outcomes from the three composition conditions: A (visual-only), B (audio-only), and C (audiovisual), combining descriptive summaries of the Likert items (see Table 2; Figs 1–4), parameter-choice counts (see Fig 6), and thematic synthesis of free-text reflections. Given the small sample (N=9), quantitative results are reported descriptively as medians [interquartile range] and paired plots are used to highlight within-participant patterns.

4.1 Participants and data completeness (Table 1)

Nine participants completed all three blocks (N=9; 27/27 blocks; 0% missing quantitative items). All sessions were solo and block order was counterbalanced. Musical experience ranged from none to advanced/professional; theory familiarity ranged from none to high. One participant reported red–green colour deficiency, and one used speakers rather than headphones. Participant background metadata are summarised in Table 1.

4.2 Condition preference, perceived intermediality, and mismatch (Fig 1)

End-of-session rankings show a strong preference for the audiovisual condition (C): 8/9 participants ranked C as their top condition (with 1/9 preferring B; 0/9 preferring A). Participants also most frequently selected C as “most intermedial” (7 selected C; 1 selected B; 1 response did not provide a single-condition selection and was coded as unsure/blank). In contrast, participants most often identified the audio-only condition (B) as producing the “biggest mismatch” (6/9), compared with A (2/9) and C (1/9). These distributions are shown in Fig 1.

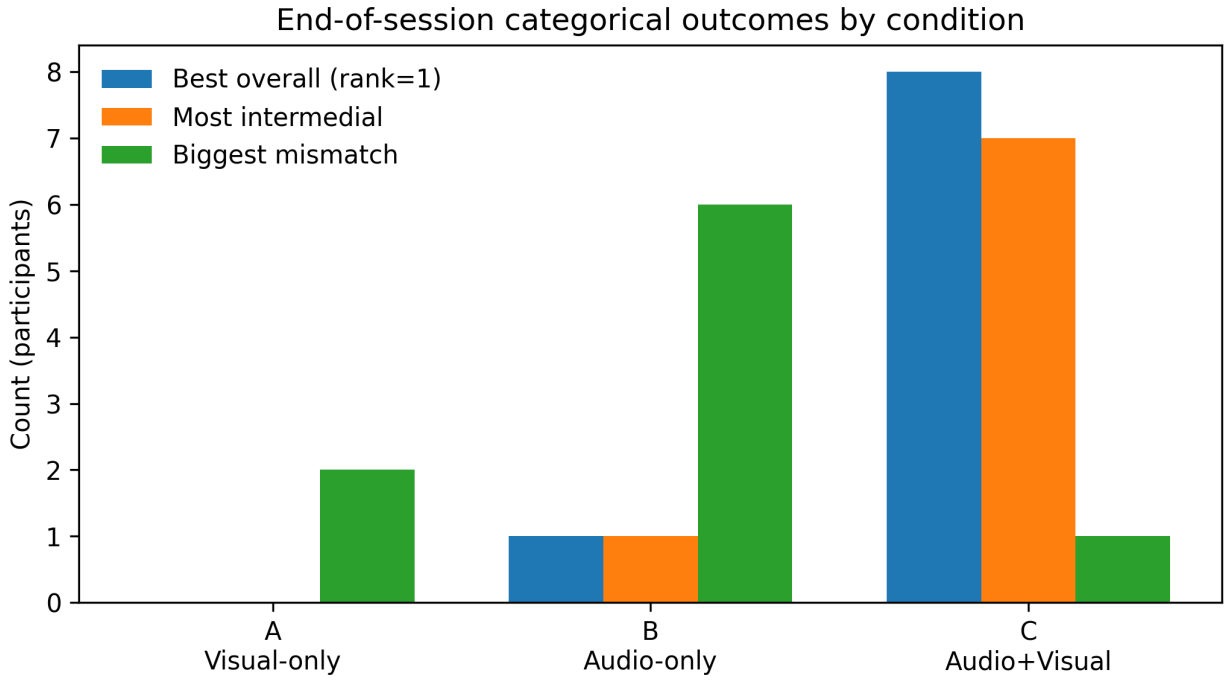

Figure 1. End-of-session counts for preference, most intermedial, and biggest mismatch.

A montage of visual outcomes across participants and conditions is provided in Figure X.

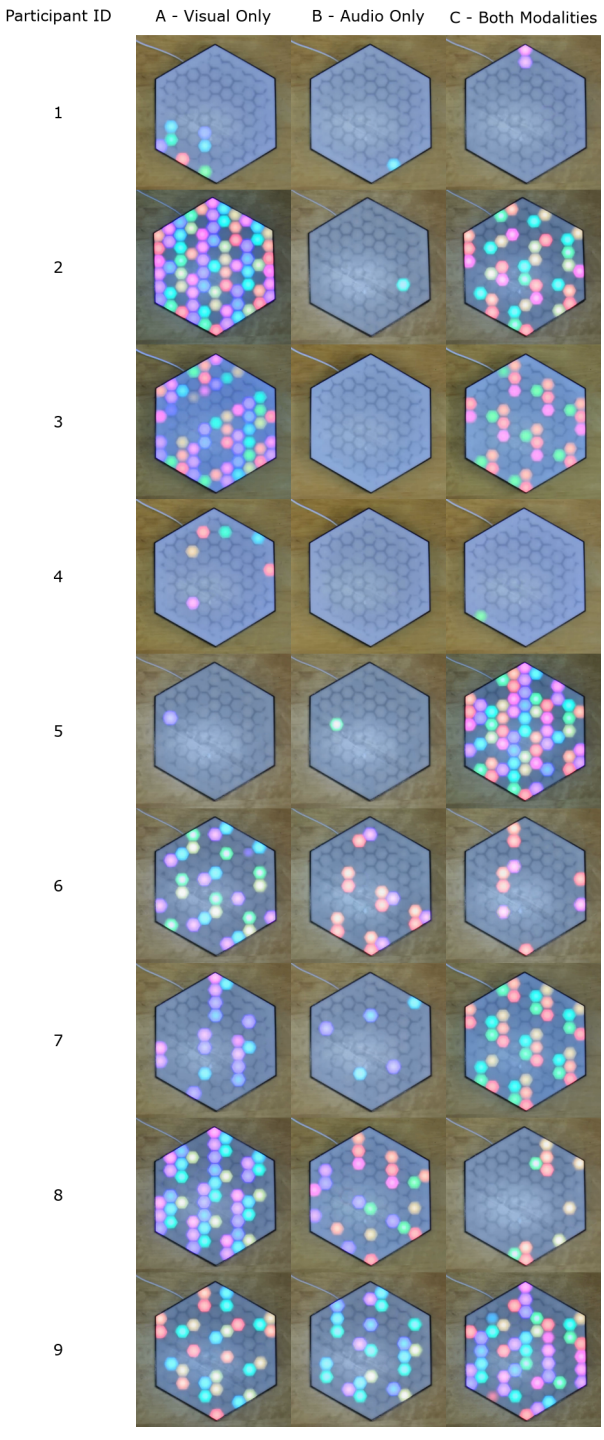

Figure X. Video stills montage of participant outputs across conditions (A/B/C).

Taken together with the free-text accounts, these outcomes suggest that “intermediality” is not treated by participants as a narrow judgement of mapping correctness. Instead, it functions as an experiential judgement about whether the coupled system supports (i) intention formation, (ii) monitoring and causal inference during interaction, and (iii) retrospective sense-making during replay.

4.3 Compositional strategies under constraint: density, arcs, loops, and exploratory testing

Across conditions, participants described compositional approaches that clustered into five recurring strategy families:

(i) density shaping (e.g. filling the grid and thinning it),

(ii) arc/narrative forms (build-up toward chaos followed by release),

(iii) loop-based structuring to establish predictability

(iv) harmony/scale manipulation (sometimes using colour similarity as a proxy)

(v) rapid A/B testing of parameters to learn local cause–effect relations.

However, the emphasis of these strategies tended to shift with modality constraint.

In visual-only (A), participants commonly treated the panel as a dynamic visual field and composed by covering the grid, sculpting density, and exploring emergent pattern families. Consistent with these accounts, perceived steerability (A3) was highest in A (6.00 [4.00, 6.00]; Table 2; Fig 2), suggesting that visual feedback supported a confident action–evaluation loop even when audio was unavailable.

In audio-only (B), strategies were more often framed in texture, pacing, repetition, and risk management. Looping was described as a stabilising container to prevent drift into uncontrolled noise. Quantitatively, steerability dropped in B (4.00 [3.00, 5.00]; Table 2; Fig 2), aligning with descriptions of reduced state visibility and uncertainty about what the system was doing.

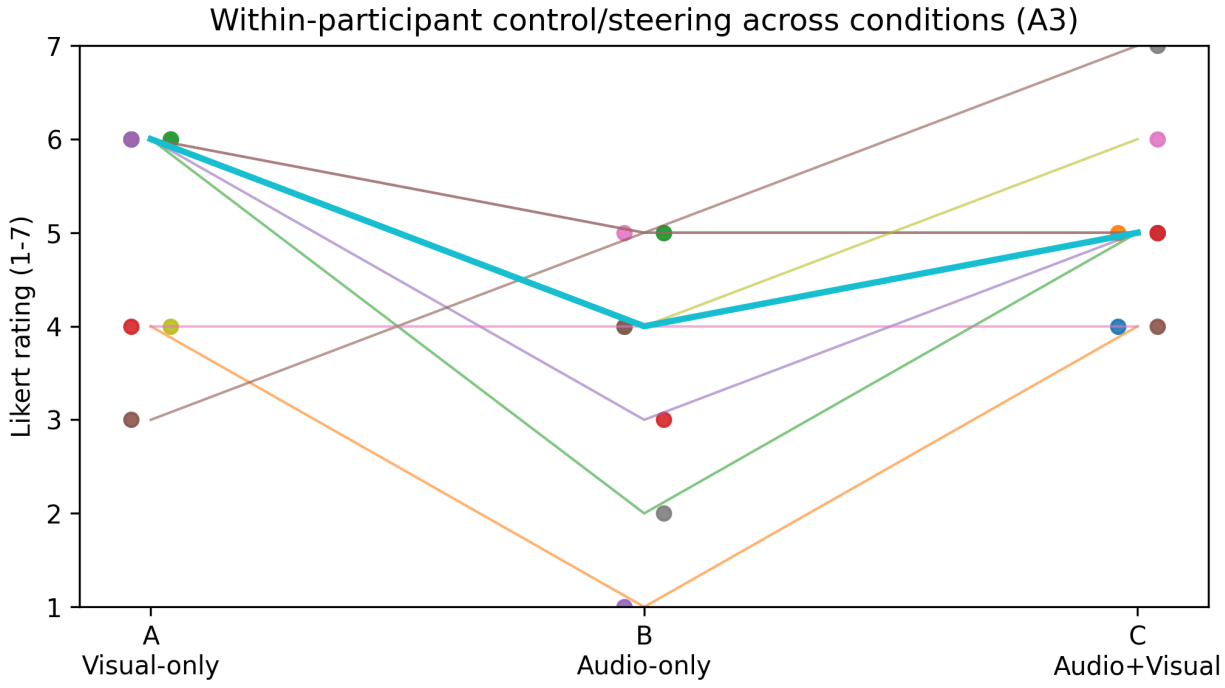

Figure 2. Paired steerability (A3) by condition.

In audiovisual (C), participants frequently alternating more deliberately between density control, harmonic selection, and temporal shaping. Satisfaction (A1) was highest in C (6.00 [5.00, 6.00]; Table 2) and useful surprise (A5) also increased (6.00 [5.00, 6.00]), supporting accounts of C as enabling both intention and productive emergence. At the same time, some participants noted the workload of attempting to optimise sound and light simultaneously; frustrating unpredictability (A6) was slightly higher in C (4.00 [3.00, 4.00]) than in A (3.00 [2.00, 4.00]) or B (3.00 [2.00, 5.00]) (Table 2).

4.4 The reveal as a site of intermedial rebinding and reinterpretation (Fig 3; Table 2)

A recurring phenomenon was the reveal (immediate replay with both modalities enabled after composing under constraint) acting as a moment of retrospective binding and reinterpretation. Participants frequently described surprise, delight, or a re-evaluation of what they believed they had controlled; several accounts described replay as disclosing latent structure in the shared generative substrate.

This is supported by item B1 (“two views of the same underlying process”), which was higher after audio-only and audiovisual composition than after visual-only composition

(A: 4.00 [3.00, 6.00]; B: 6.00 [6.00, 7.00]; C: 6.00 [6.00, 6.00]; Table 2), summarised in Fig 3. In other words, participants were most likely to judge the coupled replay as "the same process" when sound had been salient during composition (B) or when both modalities were available (C), whereas composing visually produced weaker same-process judgements for some participants.

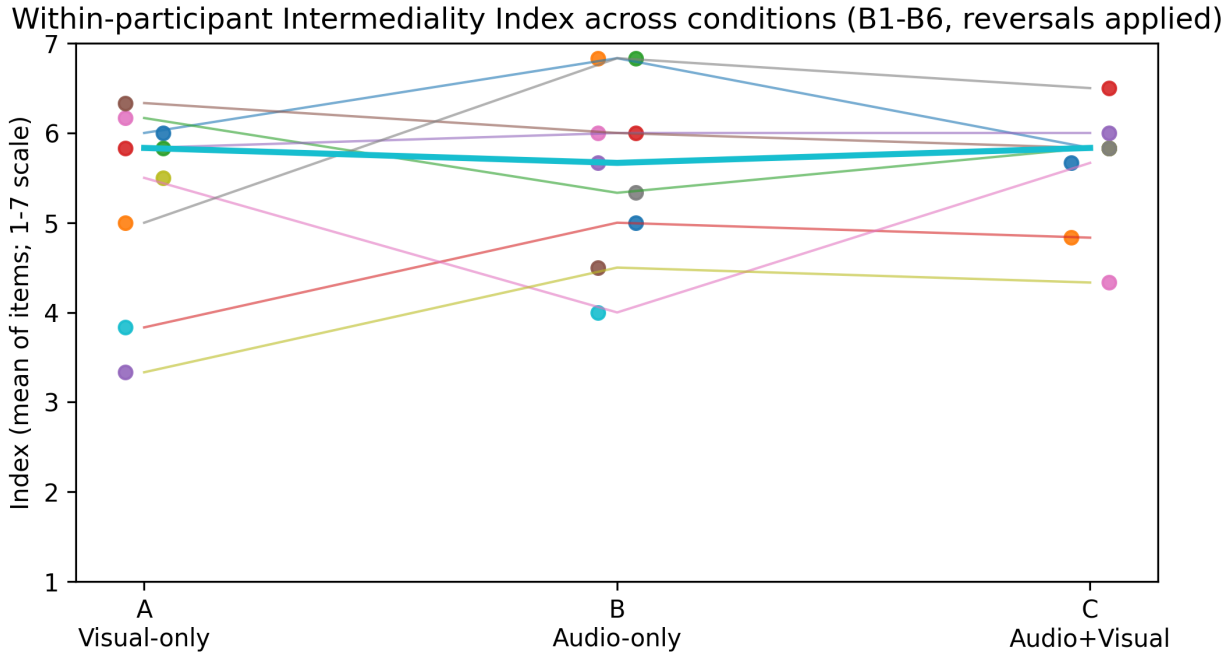

Figure 3. Paired intermediality index by condition.

These results also clarify why B was often selected as the "biggest mismatch" (Fig 1) without necessarily being experienced negatively: in multiple accounts, mismatch operated as a productive discontinuity, where replay exceeded expectations and functioned as a calibration signal that extended the participant's causal model of the system.

4.5 Attentional hierarchy and "media equality" as a negotiated achievement (Table 2)

Participants repeatedly articulated an attentional hierarchy between modalities. Some reported that sound dominated experience and that audio needed to be removed before they could prioritise visual composition; others described the inverse, noting the difficulty of controlling audio-only states despite later recognising more structure on replay.

Quantitatively, “balanced modalities” (B2) showed similar medians across conditions (A: 6.00 [2.00, 6.00]; B: 6.00 [3.00, 6.00]; C: 6.00 [3.00, 6.00]; Table 2), but with wider dispersion in A. This supports a core qualitative theme: “media equality” is not an intrinsic property of the instrument; it is negotiated through attentional dominance, familiarity, and the availability of stabilising cues (notably repetition and looping).

Perceptual overload (B6) remained low across conditions (A: 2.00 [1.00, 2.00]; B: 2.00 [2.00, 3.00]; C: 2.00 [2.00, 2.00]; Table 2; optional Fig 4). Although participants sometimes used terms such as “chaos” or “busy,” overload did not emerge as the dominant failure mode at group level; mismatch and interpretability were more diagnostic, particularly in the audio-only condition.

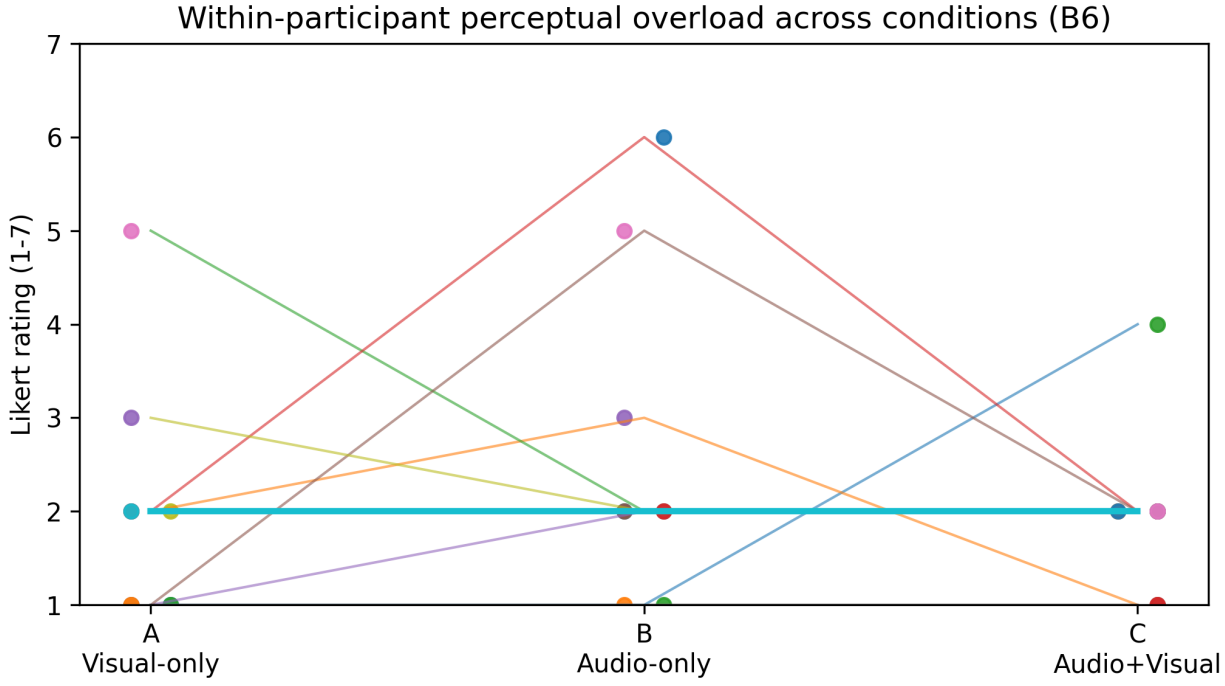

Figure 4. Paired perceptual overload (B6) by condition.

4.6 Intermedial interference as cue conflict: temporal mismatch, density mismatch, and perceptual masking

When interference was described as disruptive, it was typically framed as cue conflict, where one modality implied temporal or causal relations that the other did not support.

The most concrete conflict concerned duration cues: participants noted that light persistence could be misleading relative to note length, weakening temporal correspondence and cross-modal prediction. A second conflict concerned density cues: sparse visuals were sometimes perceived as sonically busy, while visually dense states could be perceived as less complex than expected. A third issue was perceptual masking introduced by timbre and octave stacking, described as collapsing multiple events into a single percept and reducing the auditability of harmonic relationships. Collectively, these reports align with the concentration of “biggest mismatch” selections in B (Fig 1) and lower steerability in B (Fig 2): when cues are withheld or unreliable, participants’ internal causal model becomes less stable during action.

4.7 Agency, autonomy, and co-creative negotiation with an autonomous substrate (Table 2)

Across conditions, participants framed interaction less as direct control and more as negotiated agency, emphasising the value of steerability coexisting with surprise. One participant explicitly questioned authorship (“whether I performed or the Tonnetz performed”), capturing a recurring ambiguity in how creative credit is distributed between human and system.

This theme is consistent with autonomy ratings (B10), which were moderate in A and B (both 4.00 [3.00, 5.00]) and slightly higher in C (5.00 [2.00, 5.00]; Table 2). Qualitatively, audiovisual composition often supported a shift toward a co-creative stance, acceptance and collaboration with the instrument, particularly after participants had experienced both constrained conditions and developed a more stable causal model.

4.8 Parameter strategy and interpretability: what participants used to steer (Fig 6)

Self-reported “most influential parameters” show both a stable core and a modality-dependent shift (Fig 6). Rate was the dominant control across all conditions (A: 8 selections; B: 7; C: 6), consistent with participants using global dynamism to structure arcs and transitions. Scale exhibited the clearest modality dependence: it was cited far more often when sound was available during composition (B: 7; C: 8) than in visual-only composition (A: 3), suggesting that harmonic admissibility becomes a primary steering dimension when participants can audition outcomes in real time. Other parameters (Life Length; neighbour and population thresholds; loop length/on-off) were cited at moderate frequency across conditions, supporting qualitative descriptions of “constraint tuning” to locate stable regimes and explore local variation.

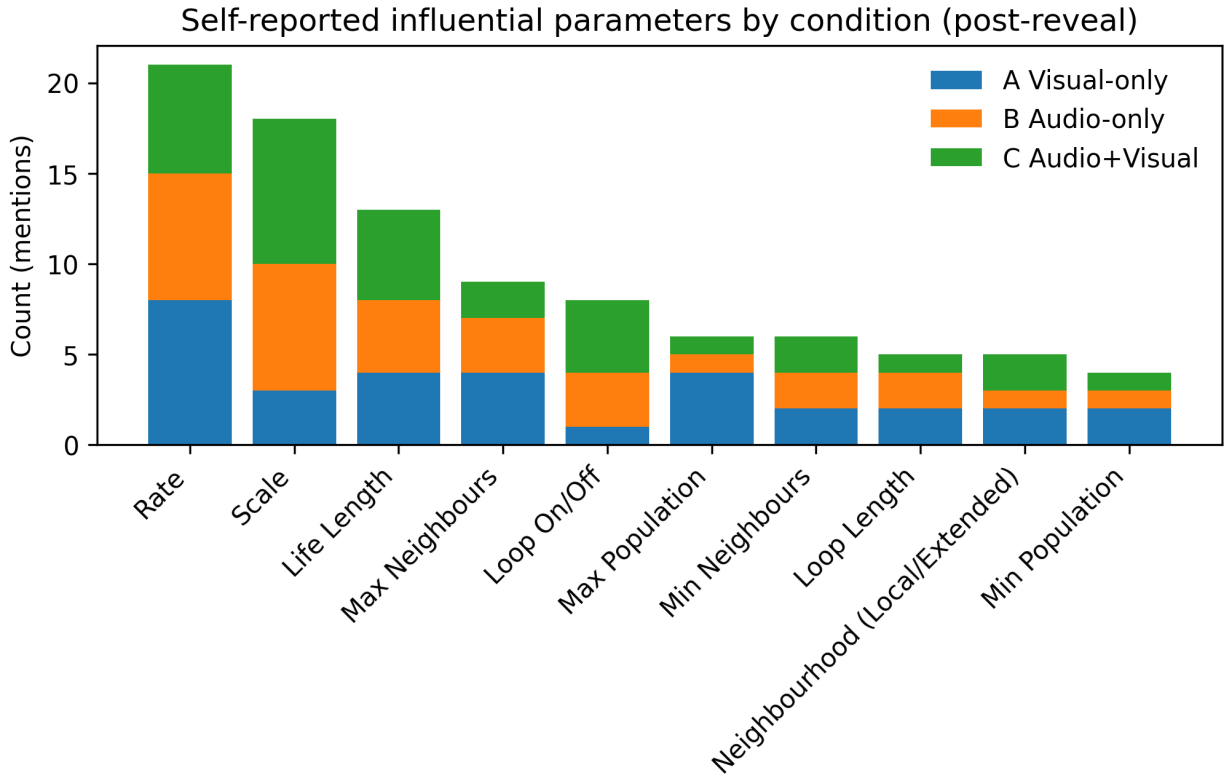

Figure 6. Parameter influence counts by condition.

4.9 Participant-led design requirements: improving legibility, prediction, and expressive articulation

Participants produced convergent, instrument-actionable requirements that map directly onto the mechanisms above: clearer loop and reset affordances; improved parameter semantics and scaling (including counter-intuitive rate behaviour); reduced interface fragmentation; improved colour fidelity and discriminability; better temporal correspondence between note duration and light persistence (or an intentional decay model communicating release); and richer articulation/timbre to reduce perceptual masking and improve event discriminability. These requirements connect to the results in a direct way: where cue reliability and parameter semantics were weak, participants reported mismatch and reduced controllability; where stable cues were available (especially repetition/looping), participants reported safety, predictability, and more intentional structuring.

4.10 Summary

Across conditions, intermedial relations in this instrument were experienced as a negotiation between cue reliability, attentional dominance, and an evolving causal model of a shared generative substrate. Constraint conditions did not simply degrade experience; they reconfigured how participants formed intention, monitored outcomes, and attributed agency; sometimes producing negative mismatch (loss of legibility) and sometimes producing positive mismatch (reveal-driven rebinding and delight).

5 Discussion

5.1 Media equality as negotiated attention rather than guaranteed parity

A central motivation of this paper was to treat “media equality” as an empirical question rather than an a priori aesthetic claim. Although the instrument enforces *causal parity:* sound and light are generated as parallel projections of a shared evolving state, participants’ reports and rankings show that causal parity does not automatically yield *perceptual or interpretive parity*. Instead, equality emerged as a negotiated achievement shaped by attentional dominance, the reliability of cross-modal cues, and the participant’s evolving causal model of the system. The strong preference for the audiovisual condition is therefore not evidence that unrestricted access is intrinsically “better,” but that simultaneous access better supports the full intention-monitoring-interpretation loop: participants can form intentions with reference to both media, test local hypotheses during interaction, and confirm or revise causal inferences in real time. In this sense, the instrument’s intermedial promise is realised not merely through coupling, but through the conditions that allow participants to *use* coupling as a resource for sense-making.

5.2 Intermedial interference as cue conflict, and mismatch as productive discontinuity

Participants’ accounts clarify intermedial interference as primarily a matter of *cue conflict*: one modality implies temporal, structural, or causal relations that the other does not consistently support. Three conflicts were especially salient. First, duration cues: light persistence can imply sustained events even when notes have released, weakening temporal correspondence and reducing predictive power. Second, density cues: visually sparse states can produce sonically busy textures (e.g., through octave stacking or cascades), while visually dense states can sound less complex than expected, destabilising the common heuristic that “more lights equals more sound.” Third, perceptual masking: timbral simplicity and stacked pitch relations can collapse multiple events into a single percept, reducing harmonic auditability and thereby undermining the cross-modal legibility required for confident steering, particularly in audio-only work. These conflicts align with lower steerability in audio-only composition and with participants’ frequent designation of audio-only as the “biggest mismatch.”

However, mismatch was not uniformly negative. A notable finding is that the reveal/replay frequently converted mismatch into a *productive discontinuity*: participants described surprise and delight when the reinstated audiovisual artefact appeared more

coherent than the compositional experience that produced it. This suggests an important nuance for interference-focused accounts of intermediality: destructive interference is not synonymous with surprise or divergence. Instead, the valence of mismatch depends on whether divergence yields interpretive dead-ends (loss of legibility) or interpretive openings (a rewarded re-binding that extends the participant's causal model). In a coupled generative instrument, productive mismatch may function as a learning mechanism, calibrating expectations and expanding the space of viable compositional strategies.

5.3 The reveal as a site of binding and retrospective causality attribution

Methodologically, the "reveal" protocol; composing under constrained access, then immediately replaying with both modalities present, proved to be a high-yield probe of intermedial construction. It exposed how participants form partial world-models from one stream, how they anticipate the withheld stream, and how they reconcile or reinterpret relationships once full audiovisual information becomes available. The elevated same-process judgements (B1) after audio-only and audiovisual work suggest that auditory salience during action supports stronger retrospective binding, whereas visual-only composition can yield visually coherent outcomes that do not always retrospectively read as "the same process" once sound is introduced. More broadly, the reveal highlights how intermediality is not only a property of the artefact but also of the *interpretive process*: participants revise causal stories about what "drives" what, sometimes reassigning agency from self to system or vice versa. This is consistent with the broader audiovisual literature that treats audio-vision as an integrated perceptual mode rather than a simple addition of channels, and it supports the use of replay-based reflection as a practical evaluation mechanism for practice-based intermedial research.

5.4 Design implications for state-space projection instruments

Because interference in this system manifests as cue reliability problems, design improvements can be framed as interventions on cue structure rather than superficial "polish." Several participant-led requirements converge into a small set of high-impact design directions:

- **Temporal correspondence:** introduce a principled light decay model tied to note release (or a matched envelope), so that duration cues communicate articulation rather than persistence.
- **Event discriminability:** adopt richer articulation (e.g., envelopes, timbral differentiation, controlled voicing) to prevent octave stacking and pure tones from collapsing distinct events into indistinguishable percepts.

- **Mapping legibility:** improve colour fidelity and discriminability, and consider persistent reference cues (e.g. inactive-note indicators) so that harmonic planning does not rely solely on transient flashes.
- **Constraint transparency:** refine loop semantics, add explicit reset/preset affordances, and clarify parameter units/scaling, particularly where control labels (e.g. “rate”) behave counter-intuitively.
- **Interface ergonomics:** reduce interface fragmentation and support faster “micro-corrections,” since intermedial steering relies on rapid hypothesis testing rather than one-shot parameter setting.

These recommendations are not generic usability claims. They follow directly from the observed mechanism: where cross-modal cues were stable and interpretable (often via looping), participants reported safety and intention; where cues were ambiguous or masking occurred, participants reported mismatch, reduced controllability, and weaker causal confidence.

5.5 Limits and scope of claims

The study is intentionally small and practice-proximate: convenience sampling, short sessions, a single audio synthesis configuration, and a limited control parameter subset. Consequently, the paper does not claim general population effects or stable effect sizes. Instead, its contribution is to demonstrate that intermedial equality and interference can be operationalised and evaluated within realistic artistic development timelines using a within-subject modality manipulation, baseline captures, saved presets for traceability, and immediate replay-based reflection. The results therefore support design-sensitive inferences about this instrument and this mapping regime, while motivating larger studies that can test generality across synthesis choices, visual mappings, and user populations.

5.6 Exploratory dyad trials

Two dyad-only pilot trials tested asymmetric modality access: one participant monitored audio via headphones while the other monitored the LED panel, with both negotiating parameter changes in real time. Although the sample is too small for inference, dyad questionnaire ratings were generally high (predominantly 5–6/7 across items), and communication notes describe an emergent division of labour in which one participant tended to drive the interface while the other provided evaluative feedback and high-level direction based on the withheld modality. Participants explicitly noted the need to alternate between affective descriptions (“feel”) and parameter-level talk, and to manage turn-taking (when to be quiet versus when to intervene), consistent with the

challenge of coordinating from complementary partial information. Across both dyad trials, participants reported moments where improvements in one modality coincided with improvements in the other, suggesting that the shared state-space can function as a common reference even under split access. In line with solo participants' reflections that dyad success likely depends on relationship and communicative fluency (rather than a single "best" modality assignment), a larger counterbalanced dyad study is a promising next step for testing whether distributed access can reliably produce coherent intermedial outcomes, and how communication strategy mediates that success.

5.7 External ratings as ongoing triangulation

A natural extension of the present evaluation is to compare participant experience against third-party judgements of the resulting artefacts. An external rating study is therefore being conducted using blinded clip presentation and standardised rating metrics; however, at the time of writing, data collection is incomplete and is reported only as progress (Appendix). The purpose of this extension is not to replace practice-based accounts with preference judgements, but to test which aspects of "intermedial success" persist as perceivable artefact properties beyond the compositional situation, and which remain primarily experiential.

6 Conclusion

This paper set out to operationalise Organised Sound's call for intermediality, specifically, the ambition that sound and moving image function as co-equal media, within a practice-based evaluation of a generative audiovisual instrument. Using the TZ5 Tonnetz cellular automata system, in which sound and LED activity are produced as parallel media projections of a shared evolving state-space, we treated intermedial interference not as an anecdotal aesthetic effect, but as a design-sensitive and evaluable property of coupled systems.

RQ1 (intermedial equality): Participants most often experienced sound and light as co-equal and "most intermedial" when composing audiovisually, but equality was not automatic; it depended on attentional dominance and the availability of stable cross-modal reference cues.

RQ2 (intermedial interference): Participants reported both constructive and disruptive interference, with disruption most often framed as cue conflict (temporal persistence, density correspondence, and event masking), and constructive interference emerging through replay-based rebinding when the withheld modality revealed latent structure.

RQ3 (agency and causality): Modality constraint systematically reshaped perceived steerability and monitoring; visual access supported more confident real-time causal inference, whereas audio-only composition amplified uncertainty during action while still allowing satisfying outcomes on reveal.

Across a within-subject protocol (visual-only, audio-only, and audiovisual composition), participants overwhelmingly preferred composing with both modalities available, yet the constrained conditions were analytically productive rather than merely degradative. They exposed how intermedial meaning-making depends on participants' ability to form intention, monitor outcomes, and update causal models during action. In particular, the replay "reveal" frequently operated as a site of retrospective binding: participants reported that the reinstated audiovisual artefact could feel more coherent than the compositional experience that produced it, reframing mismatch as either a failure of cue legibility or a productive discontinuity that extended understanding of the underlying process.

A central empirical implication is that causal coupling alone does not guarantee perceptual or interpretive equality. Instead, "media equality" emerged as a negotiated achievement shaped by attentional dominance, learned inference, and, crucially, **cue reliability** across modalities. When interference was described as disruptive, it most often took the form of cue conflict: temporal mismatches between light persistence and note duration, density mismatches between visual activity and perceived sonic

complexity, and perceptual masking that reduced event discriminability and weakened cross-modal auditability. These observations translate directly into design requirements for state-space projection instruments: align temporal cues (e.g., decay tied to note release), improve event discriminability (e.g., articulation and timbral differentiation), strengthen mapping legibility (e.g., colour fidelity and reference cues), and clarify loop/reset and parameter semantics to stabilise prediction–action–evaluation cycles.

Methodologically, the study demonstrates an evaluation approach suitable for practice-based intermedial research under time constraints: modality-access manipulation, immediate replay-based reflection, baseline system-alone captures, and traceable artefact saving (audio/video and end-state parameter presets). The approach produces analysable evidence while retaining ecological validity for exploratory composition with autonomous generative substrates. An exploratory dyad pilot further suggests that split-modality collaboration can yield coherent outcomes via division of labour and negotiated communication, motivating a larger counterbalanced dyad study.

The work has limitations typical of practice-proximate studies: a small convenience sample, a single audio synthesis and mapping configuration, and short task durations. Consequently, the paper does not claim population-level effects. Rather, it contributes an intermedial evaluation template and a concrete account of how interference and equality manifest in a coupled Tonnetz–CA instrument, providing a basis for iterative design, replication via shared presets/firmware, and larger studies. External rater assessment is underway as a triangulation mechanism; completion status is reported separately (Appendix), and future analysis will test which dimensions of intermedial success persist as perceivable artefact properties beyond the compositional situation.

In sum, treating sound and light as parallel projections of a shared generative state-space is a viable route to intermedial composition, but achieving co-equal media relations depends on the reliability of cross-modal cues and on interaction designs that support the user's evolving causal understanding. This positions intermedial interference not only as a critical lens, but as a practical engineering and compositional problem: one that can be deliberately tuned, evaluated, and iterated within instrument design.

Acknowledgements

References

Besada, J. L., Bisesi, E., Guichaoua, C. and Andreatta, M. 2024. The Tonnetz at First Sight: Cognitive Issues of Human–Computer Interaction with Pitch Spaces. Music & Science 7: 20592043241246515. https://doi.org/10.1177/20592043241246515.

Bolter, J. D. and Grusin, R. 1999. Remediation: Understanding New Media. Cambridge, MA: MIT Press.

Chiaramonte, A. 2024. Intermedial Interference in Electroacoustic Audiovisual Composition: An Investigation into Combining, Integrating, and Fusing Sound and the Moving Image. A Portfolio of Audiovisual Compositions. PhD thesis. Bournemouth University.

Chion, M. 1994. Audio-Vision: Sound on Screen. New York: Columbia University Press.

Cohn, R. 1997. Neo-Riemannian Operations, Parsimonious Trichords, and Their "Tonnetz" Representations. Journal of Music Theory 41(1): 1–66.

Cook, N. 1998. Analysing Musical Multimedia. Oxford: Clarendon Press.

Didiot-Cook, T. 2025. Cellular Au-Tonnetz: A Unified Audio-Visual MIDI Generator Using Tonnetz, Cellular Automata, and IoT. In P. Machado, C. Johnson and I. Santos (eds.) Artificial Intelligence in Music, Sound, Art and Design. EvoMUSART 2025. Lecture Notes in Computer Science, 15611. Cham: Springer, 51–65. https://doi.org/10.1007/978-3-031-90167-6_4

Douthett, J. and Steinbach, P. 1998. Parsimonious Graphs: A Study in Parsimony, Contextual Transformations, and Modes of Limited Transposition. Journal of Music Theory 42(2): 241–63.

Elleström, L. 2010. The Modalities of Media: A Model for Understanding Intermedial Relations. In L. Elleström (ed.) Media Borders, Multimodality and Intermediality. London: Palgrave Macmillan, 11–48. https://doi.org/10.1057/9780230275201_2.

Miranda, E. R. 2007. Cellular Automata Music: From Sound Synthesis to Musical Forms. In E. R. Miranda (ed.) Evolutionary Computer Music. London: Springer, 170–93. https://doi.org/10.1007/978-1-84628-600-1_8.

O'Modhrain, S. 2011. A framework for the evaluation of digital musical instruments. Computer Music Journal 35(4): 28–42. https://doi.org/10.1162/COMJ_a_00038.

Organised Sound. 2026. Call: Electroacoustic Audiovisual Composition and Intermediality: Reconceptualising media relationships. 15 January. www.cambridge.org/

core/journals/organised-sound/announcements/call-for-papers/call-electroacoustic-audiovisual-composition-and-intermediality-reconceptualising-media-relationships (accessed 31 December 2025).

Orio, N. and Wanderley, M. M. 2002. Evaluation of input devices for musical expression: Borrowing tools from HCI. Computer Music Journal 26(3): 62–76. https://doi.org/10.1162/014892602320582981.

Rajewsky, I. O. 2005. Intermédialité, intertextualité et remédiation: une perspective littéraire sur l'intermédialité. Intermédialités 6: 43–64. https://doi.org/10.7202/1005505ar.

Spence, C. 2011. Crossmodal correspondences: A tutorial review. Attention, Perception, & Psychophysics 73: 971–95. https://doi.org/10.3758/s13414-010-0073-7.