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. This article examines intermedial interference: constructive and destructive interactions between media features, as an evaluable property of generative audiovisual instruments. We study the Cellular Au-Tonnetz 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 to control the state-space. 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 informed by parameter strategies and observed approaches (e.g. density shaping, looping, and scale manipulation). 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. Working with a novel instrument that tries to treat sound and moving image as co-equal media, this is a search for meaning that emerges from the fusion of light and sound. 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?

1.5. 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 and baseline controls.
3. An interference-focused analytic account (taxonomy and exemplars) describing how constructive and destructive intermedial interference manifests in this system, and how perceived parameter regimes and modality access relate to perceived fusion, agency and overload.

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.

2.7 Instrument under investigation

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). It outputs MIDI and light simultaneously, in response to the changing state of a cellular automata based algorithm.

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.

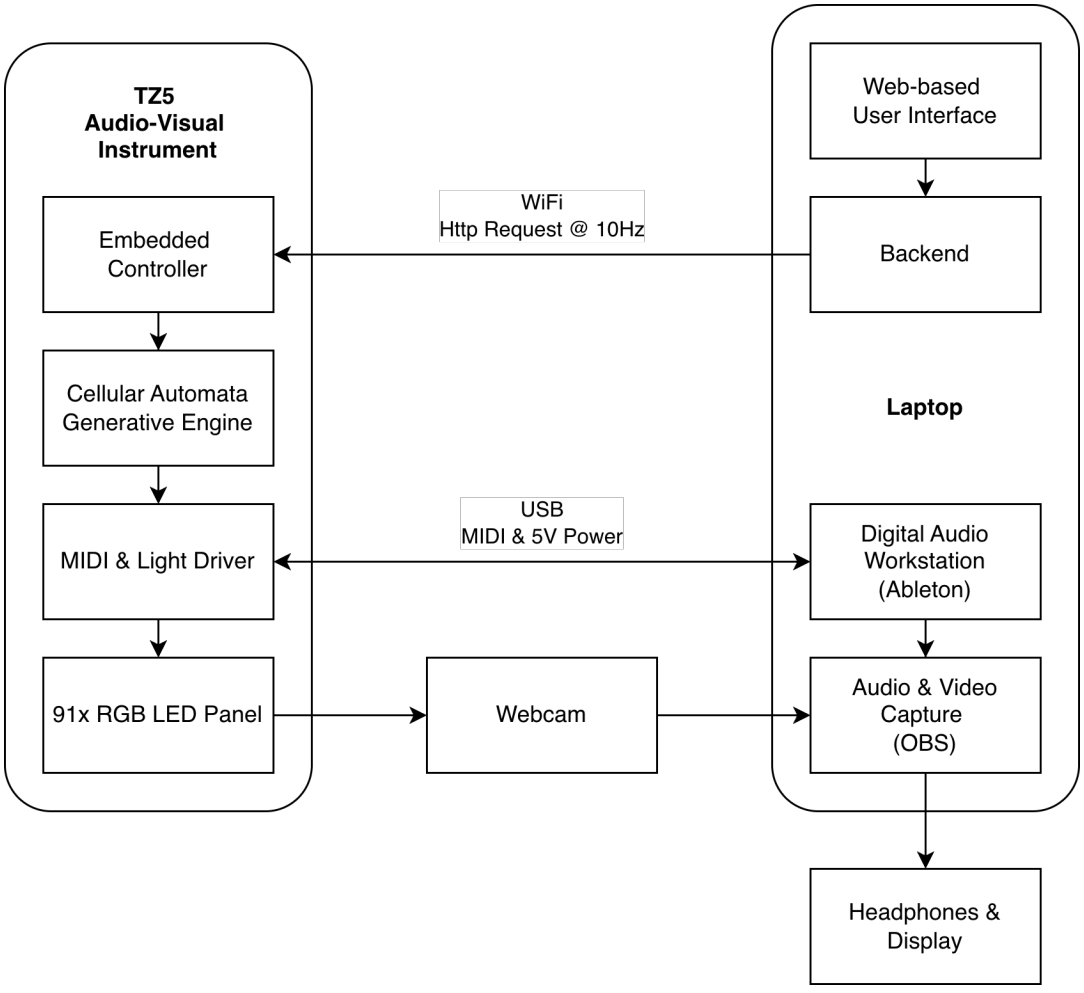

Figure 1 - System block diagram (UI → ESP32-S3 → LED panel + MIDI/audio chain).

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.

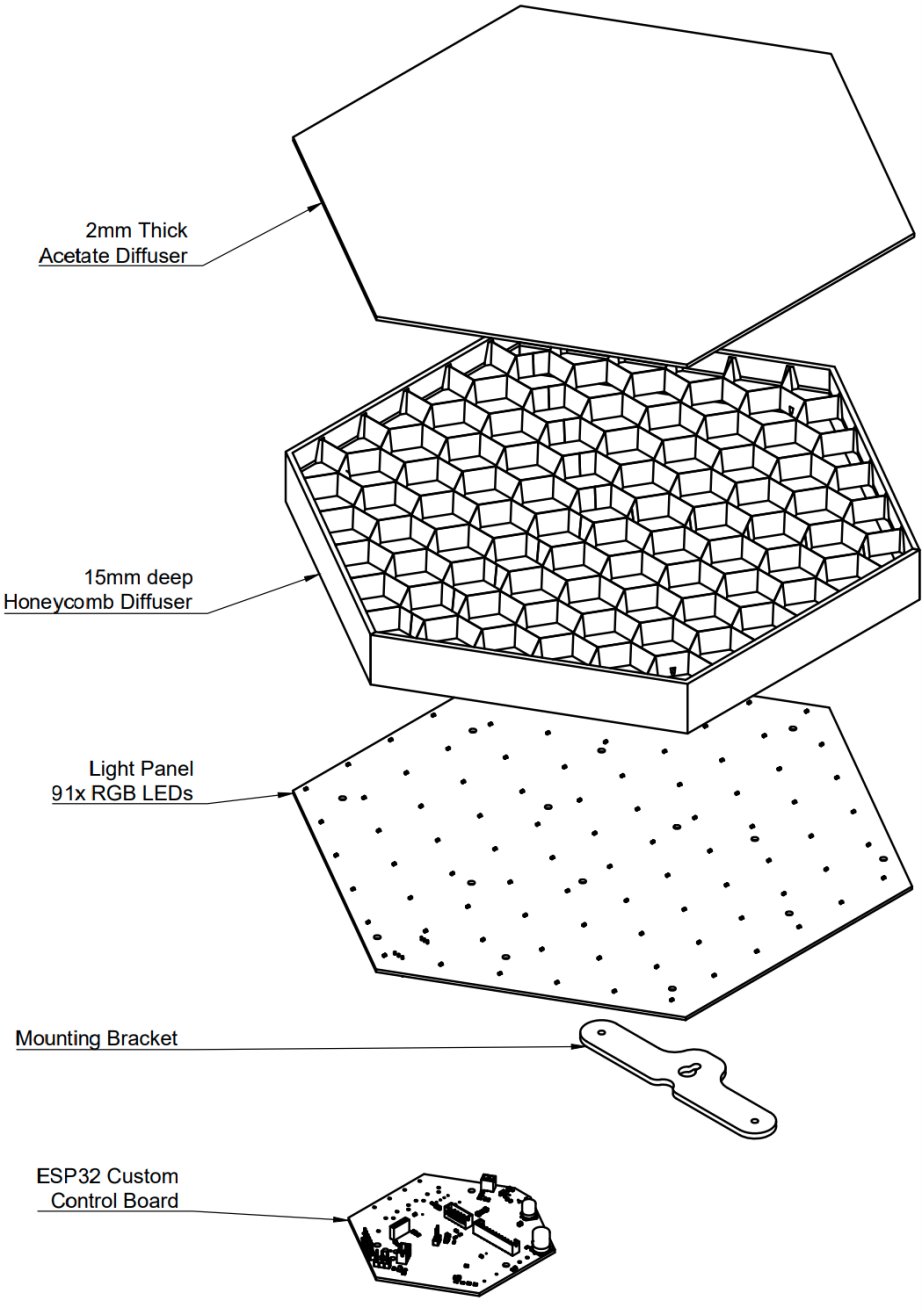

Figure 2 - Exploded view of TZ5 audiovisual instrument. Light from the 91x Red, Green & Blue (RGB) LEDs is guided by a hexagonal lattice and diffused by a 2mm sheet of acetate.

3.2.2 Coupled generation and mappings

The system implements 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.

Figure 3 shows the tonnetz pitch class and hue mapping. Hue increments around the circle of fifths, and MIDI number increments from 36 (C) to a maximum of 106 (A#).

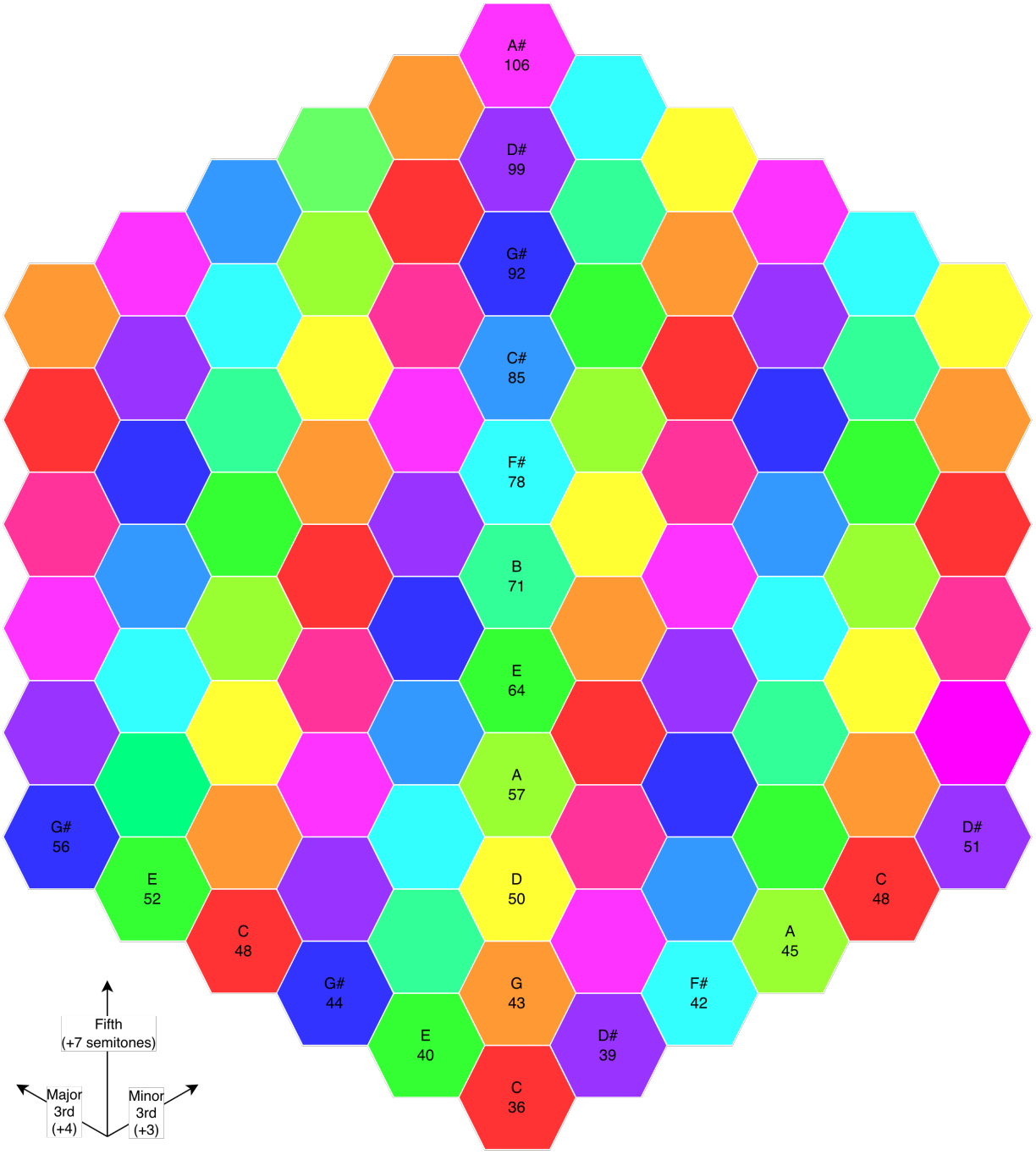

Figure 3 - Tonnetz pitch-class map used for hue mapping. MIDI numbers are given under each pitch class, and computed from the root of C;36 (bottom, centre), incrementing by +3 semitones moving right, and +4 moving left.

3.2.3 LED panel configuration

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.

3.3 Study design

3.3.1 Overview

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.

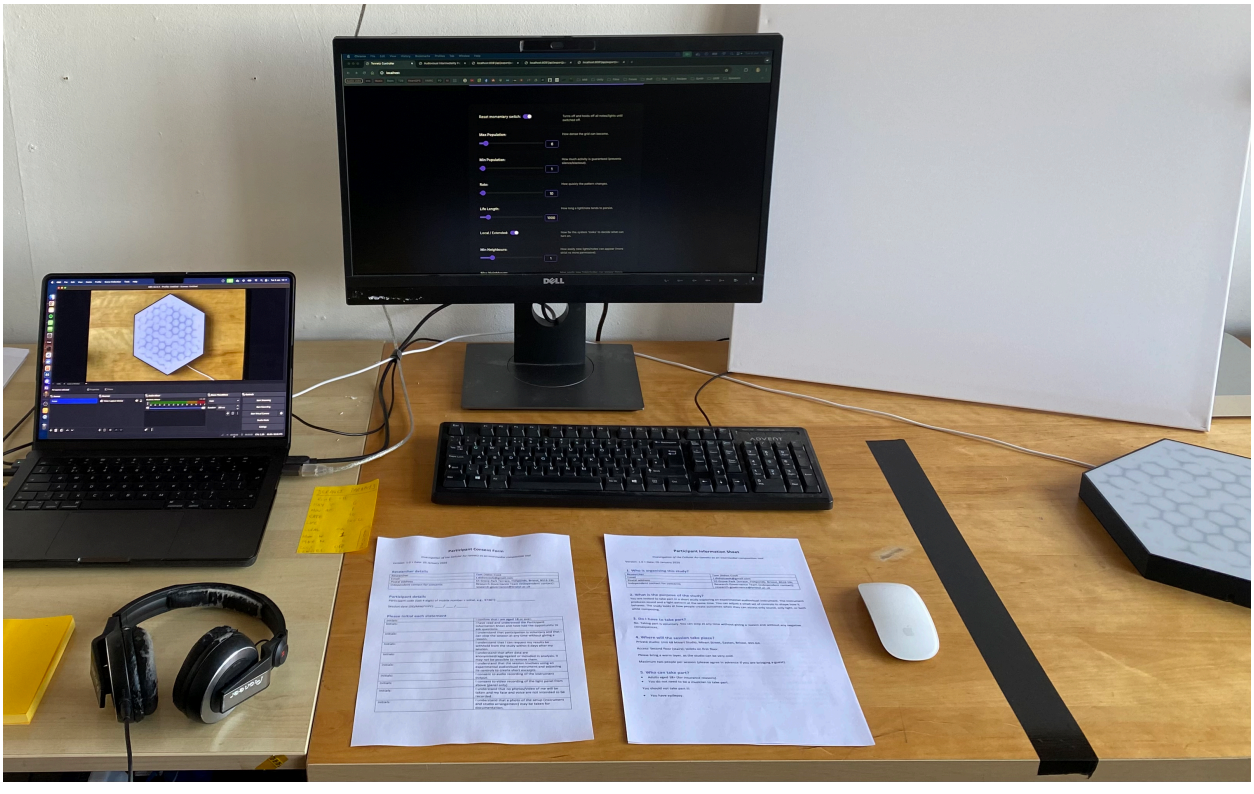

Figure 4 - Experiment setup (from left to right); Laptop hosting interface backend and recording software, Headphones, secondary display for survey and web control interface, consent form and participant information sheet, canvas for obscuring instrument, TZ5 instrument. Out of view in the top right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

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**.

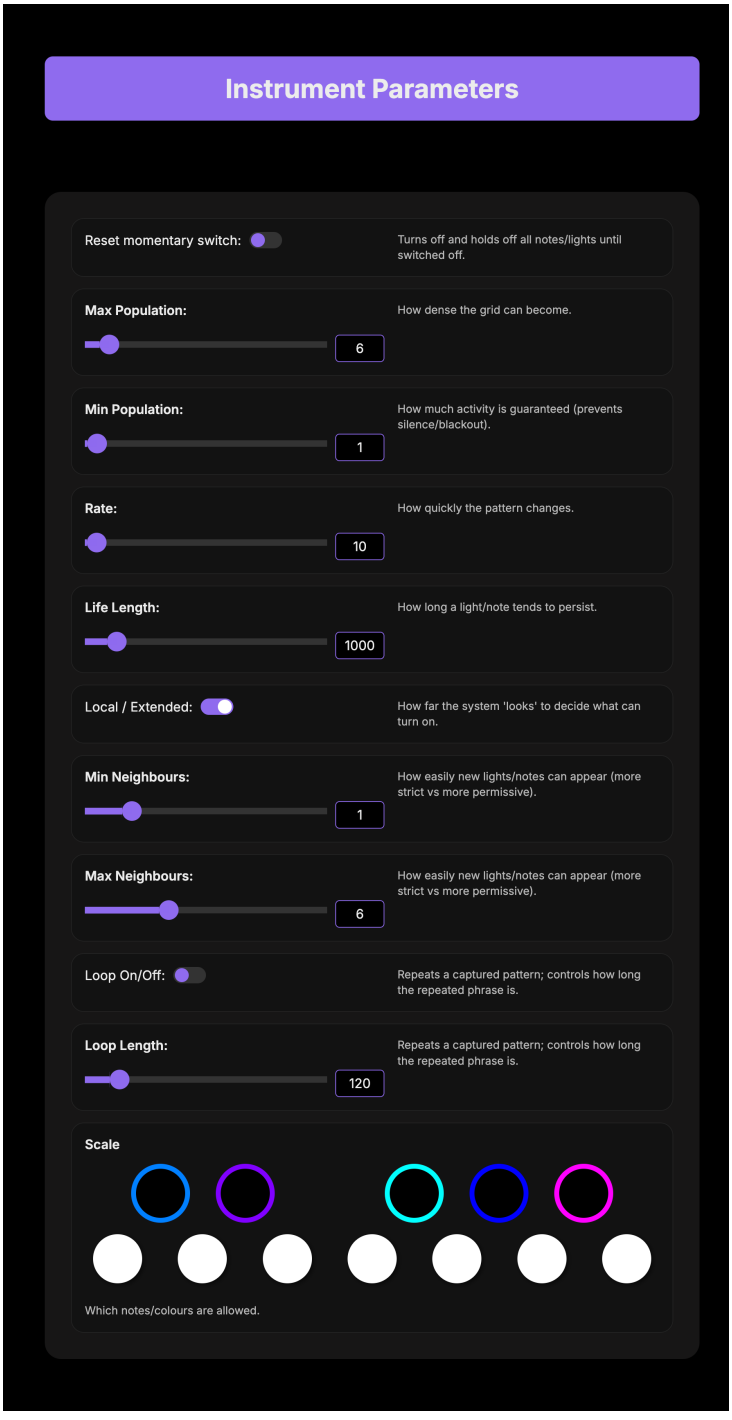

*Figure 5 - Control interface screenshot set to common starting preset (**S0**). Sliders and radio buttons allow users to steer the generative cellular-automata algorithm.*

3.3.4 Exploratory dyad protocol

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.

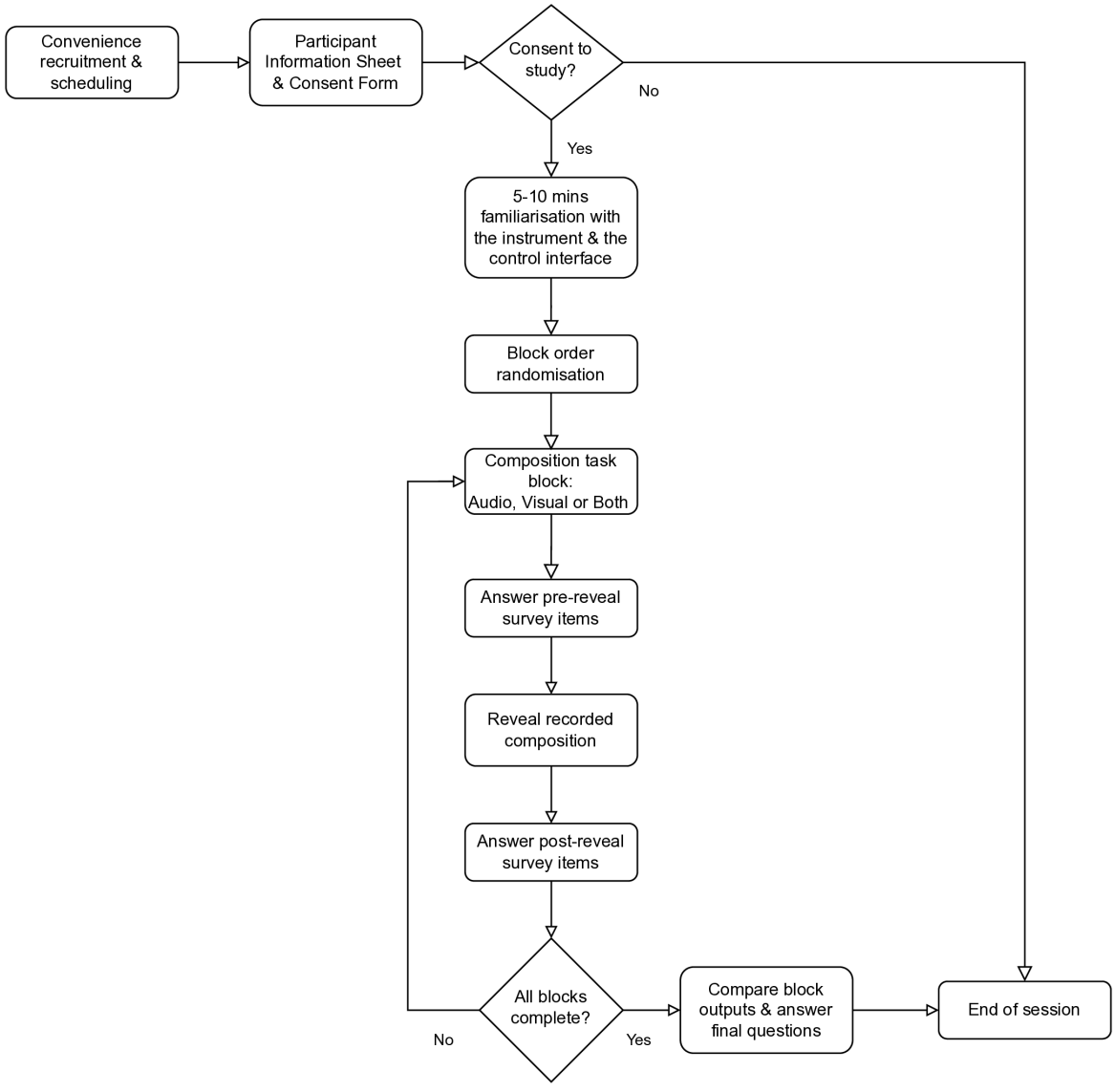

Figure 6 - Study procedure flow diagram. All 9 participants completed the full procedure.

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 to 120 seconds. Parameter changes were permitted during recording. The end-state preset for each block was saved. Users were prompted to “create something interesting” for each block.

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

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

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 and is not reported here.

3.6 Captured artefacts

The following artefacts have been captured and included as supplementary material;

- 32 video files; 1 for each block - 3 per participant, plus 3 default S0 parameter runs and 2 dyad captures.
- Part A and Part B questionnaire responses, end-of-session ranking, reflections and notebook transcriptions
- Database snapshot of all parameter changes for the 27 blocks.

4 Results: composing and interpreting intermedial relations under modality constraint

4.1 Participants and data completeness

Nine participants completed all three blocks (N=9; 27/27 blocks). 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.

ID	Age range	Musical experience	Theory familiarity	Generative experience	Tonnetz familiarity	Order
1	25–34	Some	Basic	No	No	B→A→C
2	45–54	Some	Moderate	No	No	B→C→A
3	35–44	None	None	No	Heard of it	C→B→A
4	25–34	High	High	Yes (moderate)	No	C→A→B
5	35–44	Moderate	Moderate	No	Heard of it	A→C→B
6	35–44	Some	Basic	No	No	A→B→C
7	65+	Moderate	Basic	No	No	C→A→B
8	25–34	High	High	No	No	A→B→C
9	35–44	Some	None	No	No	B→A→C

Table 1. Participant overview with anonymised IDs (1-9), background measures, and block order. Order indicates the within-participant condition sequence.

4.2 Condition preference, perceived intermediality, and mismatch

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 Figure 7.

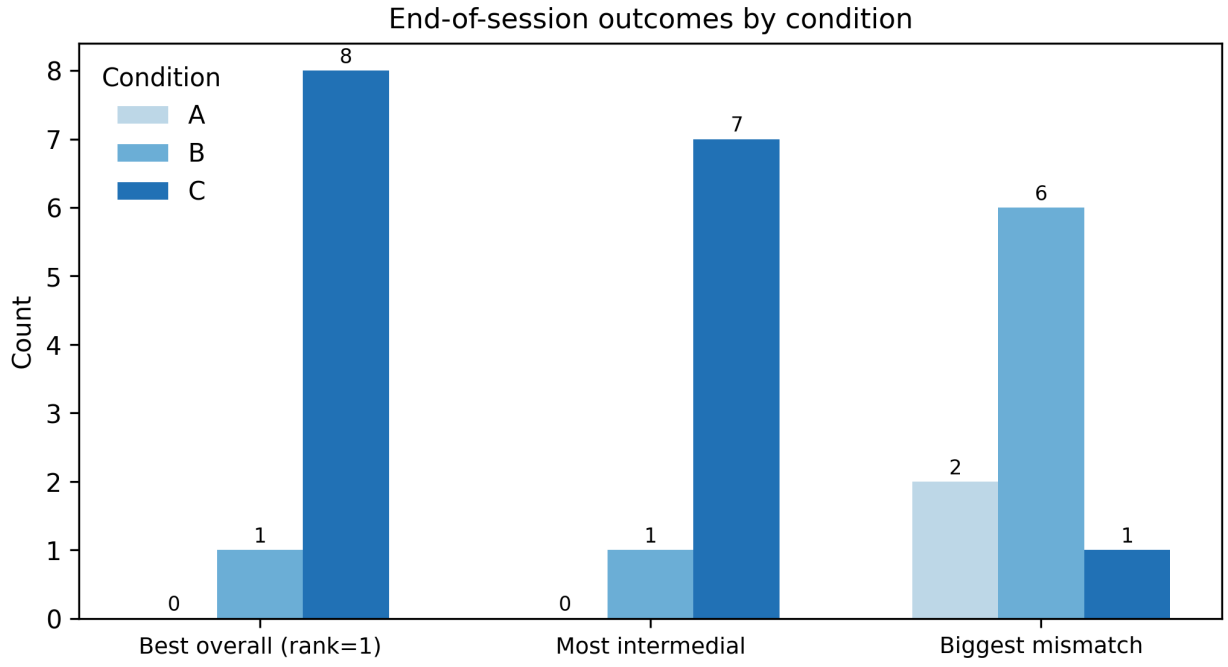

Figure 7 - Counts of best overall rank (1), most intermedial selection, and biggest mismatch by block condition (A=Visual Only, B=Audio Only, C=Audiovisual)..

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: density shaping (e.g. filling the grid and thinning it), arc/narrative forms (build-up toward chaos followed by release), loop-based structuring to establish predictability, harmony/scale manipulation (sometimes using colour similarity as a proxy), and rapid A/B testing of parameters to learn local cause–effect relations.

Figure 8 supports a key qualitative distinction: when visual feedback is available, participants more often report a stable action, evaluation loop (clear intentions, workable interaction, and steerability). In visual-only blocks, several participants described building a full field and sculpting it; “fill the whole grid,” “covering the grid,” then “reduce

intensity" or thin density to shape form over time. This aligns with the Figure 8 distributions for A2 intention clarity, A3 steerability, and A4 interface workable, which visually cluster toward agreement more than in the audio-only condition.

In contrast, in audio-only blocks, participants' strategies shift toward risk management, pacing, texture, and repetition as a stabiliser. One participant described it as "a different game" where "looping was fundamental… [a] musical container… some sense of predictability." Another noted that audio-only "definitely feels like something missing" and "didn't feel like I was really in control." This matches the Figure 8 pattern where pre-reveal items tied to controllability and clarity become more mixed in audio-only, consistent with the loss of state visibility during action.

Audiovisual blocks show a third pre-reveal profile: participants often describe actively integrating what they learned across constrained runs; "bring together what I have learnt… focus… variety," or shifting between sonic and visual priorities. In Figure 8, A1 satisfaction and A5 useful surprise visibly concentrate toward agreement in audiovisual blocks, supporting accounts that having both modalities available helps participants pursue intention while still valuing emergence and surprise. At the same time, participants explicitly report that composing with both modalities can be "overwhelming… trying to make them both look/sound good at the same time," or that different parameters "feed into the same outcome unless rules are followed." This provides a qualitative mechanism for why Figure 8 includes a strong disagreement on A6 frustrating unpredictability in audiovisual blocks: the richness of coupled feedback increases creative opportunity, but also increases the cognitive load.

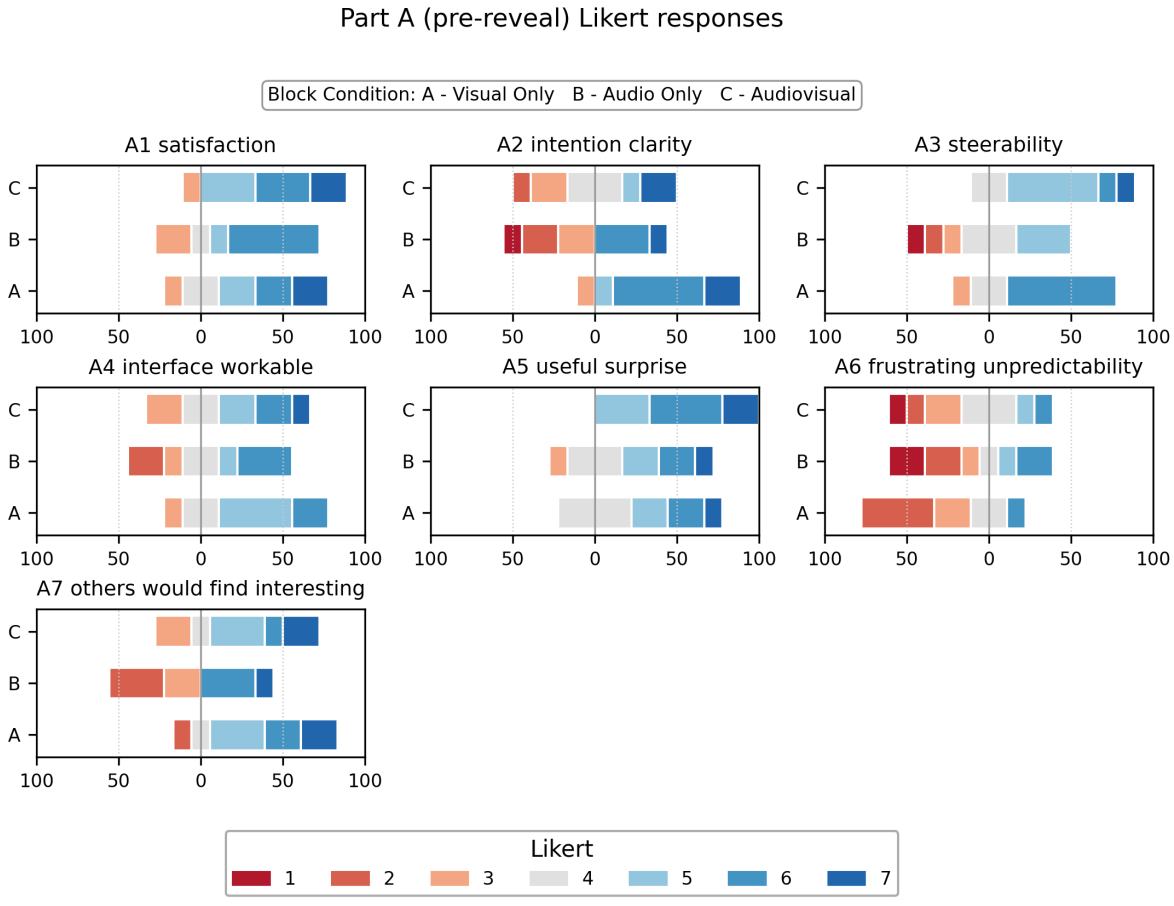

Figure 8 - Diverging stacked Likert responses for Part A (pre-reveal) items by block condition. Bars show within-condition percentages centred on neutral (4).

4.4 The reveal as a site of intermedial rebinding and reinterpretation

A recurring theme was the reveal (immediate replay with both modalities enabled after composing under one of the constraints) 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.

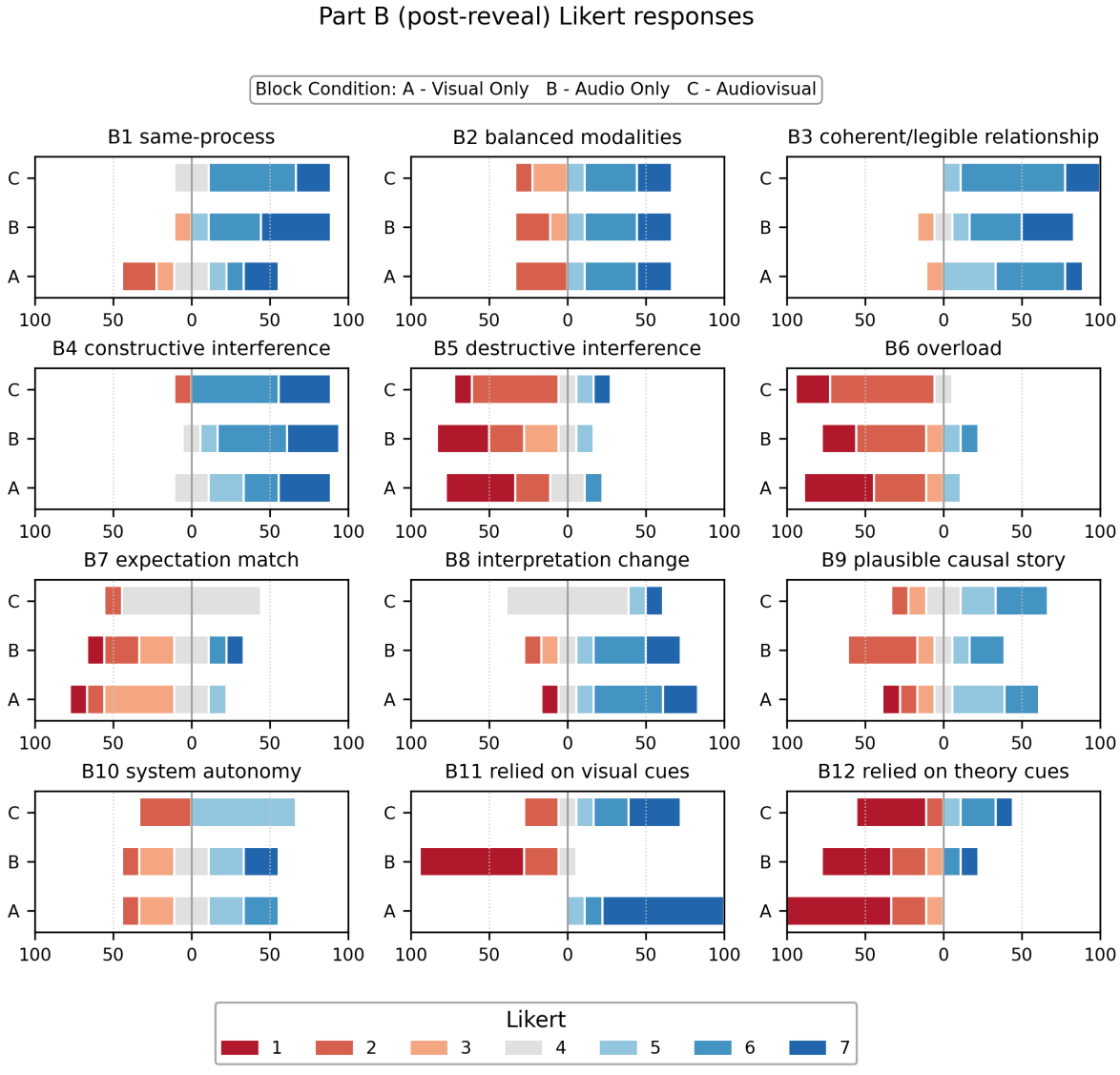

Figure 9 - Diverging stacked Likert responses for Part B (post-reveal) items by block condition. Bars show within-condition percentages centred on neutral (4).

Figure 9 supports this. After replay, participants often report that the modalities “speak to each other” and that the audiovisual result exceeded what they could predict during constrained composition. This is especially clear in audio-only blocks: one participant reports “I had no expectation of what the visuals should be, but when I saw it it was so beautiful,” framing the reveal as a positive mismatch that reshaped their understanding. Another notes the outcome was “more satisfying than I felt the recording and playing

process went," with structure becoming apparent on review. In Figure 9, this is mirrored by the response distributions for B1 sound and light felt like the same underlying process and B8 my interpretation of the output changed on reveal, which visually support the notion that replay frequently functions as a calibration moment: it either confirms that sound and light are coupled, or forces participants to revise how they believe their actions mapped onto outcomes. Qualitatively, participants describe this as discovering "more structures than expected," or as a shift from feeling tentative during action to recognising coherence after review.

4.5 Attentional hierarchy and "media equality" as a negotiated achievement

Participants repeatedly articulated an attentional hierarchy between modalities. Some reported that sound dominated their experience and that audio had to be removed before they could prioritise the visual ("the audio had to be removed completely before I could prioritise the visual"). Others described the inverse: "I definitely feel like I understand the visuals more than the audio," and noted that visual patterning revealed relationships between notes even without formal musical training. Figure 9 aligns with this theme through the cue-dependence items. The distributions for B11 relied on visual cues and B12 relied on theory cues contribute to the following:

- **Visual cues** act as the primary stabiliser for many participants' causal model. This is consistent with comments such as "without the visuals I felt less in control," and with accounts that visual pattern review "revealed more structures."
- **Theory cues** are optional scaffolding, used by some participants as a secondary strategy rather than as a prerequisite. Several participants explicitly report enjoying the experience despite limited music theory, or discovering harmony relationships through colour similarity rather than through formal knowledge.

Crucially, participants do not describe audiovisual interaction as automatically "equal" across modalities. Instead, they describe actively negotiating attention, reminding themselves to look at the panel, leaning into sound first, or using colour as a proxy for harmony after starting with theory and then migrating to the lights. This makes "media equality" best understood as an achieved practice (learned attentional switching and cue selection), rather than a fixed property of the interface.

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. Three conflicts recur across accounts:

- **Temporal mismatch:** light persistence can be misleading relative to note duration ("light carries on after note dies out"), weakening temporal correspondence.
- **Density mismatch:** sparse visuals can sound busy and vice versa ("a number of sparse visuals sounded as busy as a previously very active visual").
- **Perceptual masking:** octave stacking or timbral simplicity can collapse multiple events into a single percept ("different octaves… sounded as one"), reducing auditability of harmonic relationships.

Figure 9 provides a useful counterbalance to these failure descriptions. While cue conflicts are clearly present in the qualitative data, participants frequently describe the overall relationship as coherent ("audio and visuals aligned perfectly," "clear relationship," "coherent," "matched"). This is consistent with the distributions around constructive versus destructive interference items: rather than indicating pervasive breakdown, the Likert patterns support the interpretation that interference is typically experienced as productive, as opposed to a global failure mode.

4.7 Agency, autonomy, and co-creative negotiation with an autonomous substrate

Across conditions, participants framed interaction less as direct control and more as negotiated agency, emphasising the value of steerability coexisting with surprise. Several accounts explicitly articulate this ambiguity of authorship ("Still have a question… if I performed or the Tonnetz performed"; "we (me and instrument) is the dyad"). This co-creative framing links directly back to the Likert patterns across Part A and Part B. Pre-reveal, participants' ability to form intentions and test them is strongest when visual state is available (Figure 8), supporting a sense of meaningful steering even when outcomes remain emergent. Post-reveal, participants frequently report increased coherence and satisfaction with outcomes on review ("much better than I thought it would be"), including cases where action-time control felt limited ("more satisfying than… the process went"). Together, Figures 8 & 9 support an interaction model in which agency is not the elimination of emergence, but the achievement of *legibility sufficient for steering*: participants accept autonomy of the instrument, while relying on stabilising cues (often visual) to keep the creative loop workable.

4.8 Parameter strategy and interpretability: what participants used to steer

Self-reported “most influential parameters” show both a stable core and a modality-dependent shift (Figure 10). 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 hear 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.

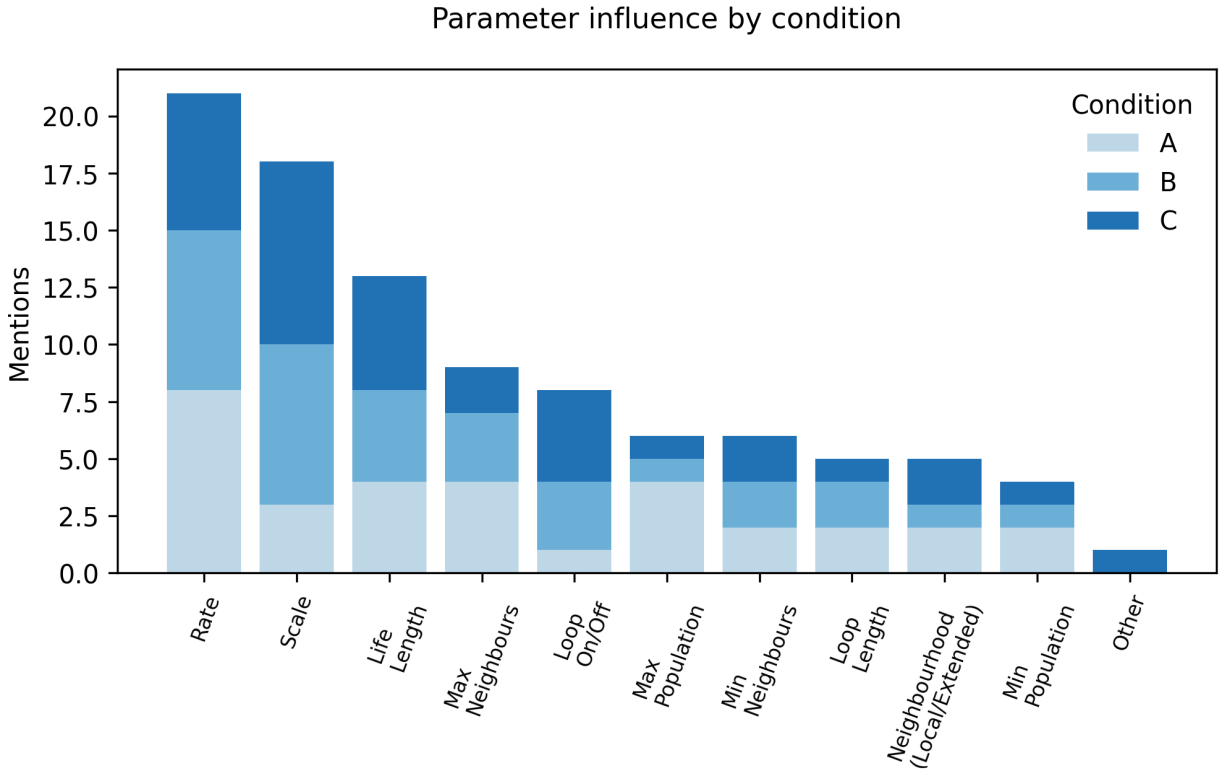

Figure 10 - Stacked bar chart of parameter influence nominations by condition.

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.

A montage of visual outcomes across participants and conditions is provided in Figure 11.

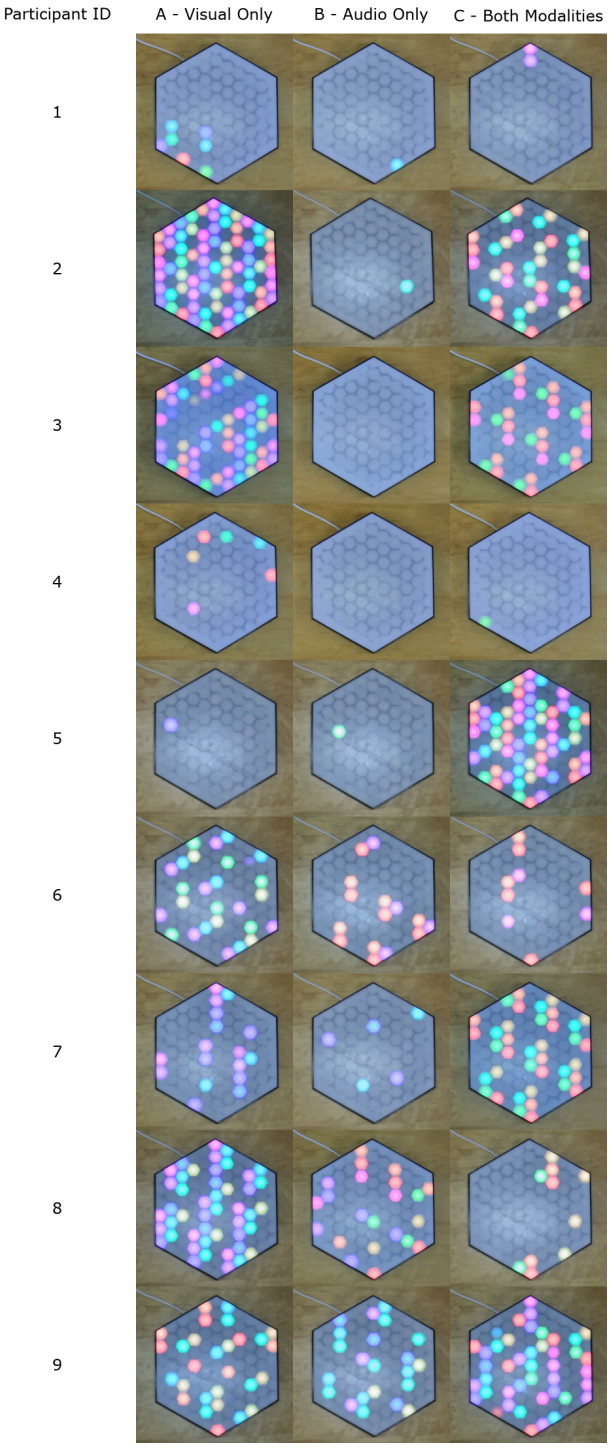

Figure 11 - Video-stills montage of participant outputs across conditions (A/B/C).

5 Discussion

5.1 Media equality as negotiated attention rather than guaranteed parity

This study treated “media equality” as an empirical question rather than an a priori aesthetic claim. Although the instrument represents sound and light as parallel projections of a shared evolving state, participants’ reports show that causal parity does not automatically yield perceptual or interpretive parity. Instead, equality emerged as a negotiated achievement shaped by attentional dominance, cue reliability, and participants’ evolving causal models. The strong preference for the audiovisual condition suggests that simultaneous access best supports the full intention–monitoring–interpretation loop: participants can form intentions with reference to both media, test local hypotheses during interaction, and refine causal inferences in real time. In this sense, the instrument’s intermedial promise is realised not merely through coupling, but through interaction conditions that let performers use coupling as a sense-making resource.

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

Participants’ accounts frame disruptive interference primarily as cue conflict: one modality implies temporal, structural, or causal relations that the other does not reliably support. Three conflicts recurred. First, duration cues: light persistence can imply sustained events after note release, weakening temporal correspondence and predictive power. Second, density cues: visually sparse states can yield sonically busy textures (e.g. octave stacking or cascades), while visually dense states may sound less complex than expected, destabilising the 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 undermining cross-modal legibility, particularly in audio-only interaction. These mechanisms align with reduced steerability under audio-only constraint and with participants’ frequent designation of audio-only as the “biggest mismatch.”

Mismatch, however, was not uniformly negative. Immediate replay with both modalities often re-valued apparent mismatch as coherent structure, indicating that interference valence depends on whether divergence produces interpretive dead-ends (loss of legibility) or rewarded re-binding (an expanded causal model). In coupled generative instruments, such productive discontinuity may function as a learning mechanism: it calibrates expectations and enlarges the space of viable compositional strategies.

5.3 Design implications for state-space projection instruments

Because interference manifested as cue reliability problems, design improvements can be framed as interventions on cue structure rather than superficial “polish.” Participant-led requirements converge into a small set of high-impact directions:

- **Temporal correspondence:** implement a principled light decay model tied to note release (or a matched envelope) so duration cues communicate articulation rather than persistence.
- **Event discriminability:** introduce richer articulation or controlled voicing (e.g. envelopes, timbral differentiation) to reduce perceptual masking and improve auditability of simultaneous events.
- **Mapping legibility:** improve colour fidelity and discriminability, and consider persistent reference cues (e.g. inactive-note indicators) so harmonic planning does not rely solely on transient flashes.
- **Constraint transparency:** refine loop semantics, add explicit reset/preset affordances, and clarify parameter units/scaling, especially where controls behave counter-intuitively.
- **Interface ergonomics:** support faster “micro-corrections,” since intermedial steering relies on rapid hypothesis testing rather than one-shot parameter setting.

5.4 Limits, scope, and extensions

The study is intentionally small and practice-proximate: convenience sampling, short sessions, a single synthesis configuration, and a constrained parameter set. Accordingly, the paper does not claim population-level effects. Its contribution is to demonstrate that intermedial equality and interference can be operationalised within realistic artistic development timelines using within-subject modality manipulation, baseline captures, saved presets for traceability, and immediate replay-based reflection.

An exploratory dyad pilot suggests that split-modality collaboration can still yield coherent outcomes via division of labour and negotiated communication, motivating a larger counterbalanced dyad study to test whether distributed access can reliably support intermedial outcomes and how communicative strategy mediates success. A blinded third-party rating study is underway but is not reported here.

6 Conclusion

This paper operationalised “media equality” and intermedial interference in a practice-based evaluation of the Cellular Au-Tonnetz, a coupled generative instrument in which sound and LED activity are parallel projections of a shared evolving state-space. Using a within-subject design (visual-only, audio-only, audiovisual) with replay-based reflection and traceable artefact capture, we treated intermediality as a design-sensitive and evaluable property of a shared generative substrate rather than a post-hoc synchronisation achievement.

RQ1 Intermedial equality: Participants most often experienced the system as “most intermedial” when composing audiovisually, but equality was not automatic; it depended on negotiated attention and the availability of stable cross-modal cues.

RQ2 intermedial interference: Interference was reported as both constructive and disruptive. Disruption was most often framed as cue conflict (duration persistence, density correspondence, perceptual masking), while constructive effects emerged when divergence supported re-binding and increased perceived coherence on replay.

RQ3 agency and causality: Modality constraint reshaped steerability and monitoring. Visual access supported more confident real-time causal inference, whereas audio-only interaction increased uncertainty during action while still sometimes yielding satisfying outcomes on review.

A central implication is that causal coupling alone does not guarantee perceptual parity: co-equal media relations depend on cue reliability and interaction designs that support performers’ evolving causal understanding. For state-space projection instruments, this motivates concrete design priorities: align temporal cues, improve event discriminability, strengthen mapping legibility, and clarify constraint and parameter semantics to stabilise prediction–action–evaluation cycles. More broadly, the study demonstrates a lightweight evaluation template for practice-based intermedial research that yields analysable evidence while preserving ecological validity for exploratory composition with autonomous generative substrates.

Acknowledgements

I would like to thank the participants who so willingly gave their time to this study, Simon Jones and Sabine Hauert for their encouragement to conduct this work, and Rosa Beesley for your never ending support.

The author used ChatGPT (OpenAI) to assist with drafting, language editing/proofing, web development for survey implementation and python scripting for results analysis. All outputs were reviewed and edited by the author, who takes full responsibility for the final manuscript and materials.

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(1): 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.

Rajewsky, I. O. 2005. Intermediality, Intertextuality, and Remediation: A Literary Perspective on Intermediality. 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.