Abstract

Electroacoustic audiovisual composition frequently aspires to parity between sound and moving image, yet compositional workflows and perceptual salience often reintroduce hierarchy, positioning one medium as structural and the other as illustrative. Responding to calls to reconceptualise 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 produced as 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 perceived coherence, novelty, agency, fusion/equality and interference, followed by immediate replay-based reflection with both modalities present. Outputs were subsequently rated by independent observers using blind presentation, and baseline system-alone captures provided a control for intrinsic generativity. We outline an analysis framework linking participant experience to perceivable artefacts and propose an interference taxonomy grounded in parameter regimes and observed compositional strategies. The study offers a practical evaluation method for practice-based intermedial research and a state-space projection framing of audiovisual composition in which sound and light operate as co-equal media.

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 prob-

lem. 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: Tonnetz cellular automata 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, neighbourrule 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?

RQ4 (Distributed authorship, exploratory): Can dyads with asymmetric modality access (one hearing, one seeing) negotiate a coherent intermedial outcome, and how is this outcome perceived by independent raters?

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. Immediately 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: 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.

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.

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. Springer Nature Link+1 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 the problem of "media equality" Intermediality is routinely invoked to signal work that happens between media rather than alongside them. Rajewsky frames intermediality as an umbrella term for "phenomena that … take place between media", emphasising that the concept is heterogeneous and best treated as a research lens rather than a single technique. epe.lac-bac.gc.ca In the Organised Sound call, this lens is explicitly tied to the ambition that sound and image operate as equals in an "Intermedial Space", where meaning is mutually shaped rather than hierarchically assigned. Cambridge University Press & Assessment A useful way to avoid vague appeals to "between-ness" is Elleström's modality model, which distinguishes media by material, sensorial, spatiotemporal and semiotic modalities, allowing intermedial relations to be discussed as specific kinds of border-crossings, couplings and transformations rather than general hybridity. Springer Link This matters for your paper because your system is not merely "sound plus lights": it is a time-based visual modality (spatiotemporal patterning in light) and an auditory modality (pitch/timbre/time), coupled through a shared generative substrate. Framing the work in modality terms helps you argue that intermediality is enacted at the level of causality (shared

state-space) and perception (co-attended streams), not simply at the level of presentation. Intermedial discourse also intersects with remediation: how media reuse, refashion, or foreground mediation itself (often framed through immediacy/hypermediacy). paas.org.pl This is relevant to Organised Sound's emphasis on the ethics and politics of remediated source materials, but it also matters more practically: generative audiovisual instruments can oscillate between transparency ("I am directly shaping the work") and foregrounded mediation ("the system is visibly doing things to me"). That oscillation is a principal driver of perceived agency and, in your terms, intermedial interference. 2.2. Audiovisual relations: from film sound to musical multimedia Audiovisual theory in film studies provides durable concepts for understanding how sound and image combine into a trans-sensory whole. Chion's "audio-vision" positions audiovisual perception as an integrated mode in which sound can shape the perceived meaning of images and vice versa, rather than being treated as separable channels. Columbia University Press This orientation aligns well with the call's intermedial space framing and supports your methodological move of testing what happens when one modality is withheld during composition and reinstated only at replay: the study becomes an experiment in how audio-vision is constructed, anticipated, and retrospectively interpreted. In musicology and multimedia analysis, Cook argues for analysing how media work together rather than presuming stable hierarchies, offering a vocabulary for relations that are not reducible to "illustration" or "accompaniment". PhilArchive This line of work is useful to you in two ways: (i) it legitimises the analytical stance that intermedial meaning arises from relations (alignment, divergence, reinforcement, contradiction), and (ii) it provides a precedent for treating audience interpretation as central evidence rather than a secondary "reception" add-on. Organised Sound has also carried related discussions in adjacent form, including attention to the visual/material conditions of sound-based practices (e.g., accounts that interrogate what "audiovisual" entails once physical artefacts and staging become central to sonic art). Cambridge University Press & Assessment That thread helps you justify LED-panel composition as squarely within the journal's remit even when the visual component is not conventional moving-image video: it is still time-based visual structuring integral to the work's signification. 2.3. Electroacoustic legacies: concrète, acousmatics and the "image" of sound A key part of the call is the suggestion that musique concrète thinking can be extended across sonic and visual domains—abstracting materials from reality while retaining traces of origin. Cambridge University Press & Assessment Your system is not concrète in the strict Schaefferian sense (it is not founded on recorded sound objects), but the logic is transferable: constraint, transformation, and the play between mimesis and abstraction can be recast as the play between recognisable musical structures (scales, harmonic neighbourhoods) and abstracted generative behaviour (CA emergence and recurrence). Within electroacoustic analysis, Smalley's spectromorphology provides an estab-

lished descriptive language grounded in perception and motion-in-time, and it is explicitly concerned with how listeners infer source and causality (e.g., via "source-bonding"). York University This is directly relevant to your evaluation: participants and raters will likely form causal stories about what is "driving" what (lights driving sound, sound driving lights, or an unseen system driving both). Those inferences are not noise; they are the mechanism by which intermedial equality or hierarchy is reconstructed in perception. Related Organised Sound work on electroacoustic imagery and space similarly underscores that electroacoustic practice frequently deals in evocation, place, and image-like listening, even when no visuals are present. ACM Digital Library This gives you a strong bridge to your "audio-only" and "visual-only" task conditions: each condition invites participants to construct a partial world-model, which is then tested against the reinstated audiovisual whole at replay. 2.4. Intermedial interference and the intermedial space The thematic call points to "Intermedial Interference" as a recent contribution: constructive and/or destructive interaction of different media features. Cambridge University Press & Assessment Chiaramonte's doctoral research develops this concept in the specific context of electroacoustic audiovisual composition, positioning intermedial practice as one in which the in-between—the interaction itself—becomes the compositional site. Bournemouth University Research Online Two aspects of Chiaramonte's framing are particularly consequential for your paper. First, it motivates an analysis that is not primarily about synchronisation as a technical achievement, but about how interaction produces perceptual and interpretive consequences (reinforcement, masking, contradiction, overload, narrative leakage, etc.). Second, it highlights an evaluative gap: if interference is phenomenological and context-sensitive, it is easy to describe but hard to compare across works or design choices. Chiaramonte explicitly notes the difficulty of measurability and the potential value of broader audience reception work. Bournemouth University Research Online Your design answers that need pragmatically by triangulating (i) compositional experience, (ii) artefact properties (captured clips and saved parameter states), and (iii) independent ratings. 2.5. Multisensory perception: synchrony, binding and crossmodal correspondences A substantial body of perceptual research indicates that audiovisual integration is shaped by temporal tolerance, stimulus type, attention, and prior exposure. Vatakis and Spence's work on audiovisual synchrony perception demonstrates that tolerance for asynchrony varies across classes of events (including musical actions), implying that "tight sync" is neither a universal requirement nor a sufficient condition for perceived integration. PubMed This is useful for interpreting your outcomes: if raters judge a clip as "fused" even when participants composed under modality deprivation, the explanation may lie less in explicit alignment and more in emergent structural correspondence or learned binding. Separately, research on crossmodal correspondences shows that people reliably associate features across senses (e.g., pitch–brightness, timbre–shape, etc.), and Spence's re-

view work is a strong anchor for positioning such correspondences as a design resource rather than an anecdotal curiosity. SAGE Journals In your dyad rationale—"people are better at recognising colour similarity than note similarity"—the more defensible claim (in an OS paper) is not that colour is inherently "easier", but that visual similarity judgments and categorical naming can support shared reference and negotiation in ways that may translate into more coherent parameter steering under time pressure. The dyad method then tests whether that negotiation advantage carries through to externally perceived audiovisual quality. 2.6. Agency, co-creativity and evaluation in generative instruments Organised Sound readers will recognise the recurring tension in generative systems between autonomy and steerability: systems can be expressive because they surprise, yet frustrating because they resist direct intention. Contemporary NIME literature increasingly treats this as a question of co-creative agency and collaborative process rather than "control accuracy". For example, ethnographic and process-focused studies of machine collaboration emphasise how musicians negotiate authorship, trust, and shared direction over extended work. NIME More general computational-creativity literature similarly foregrounds creativity as situated practice rather than an intrinsic property of an algorithm. Frontiers On evaluation specifically, the DMI community has repeatedly argued that informal assessments dominate and that structured evaluation frameworks are needed to generate comparable knowledge while respecting musical context. O'Modhrain's evaluation framework is widely cited in this space and provides a defensible reference point for articulating what your study evaluates (experience, capability, context) without collapsing into usability testing. ACM Digital Library Complementary work proposes performer- and audience-oriented evaluation approaches (including structured measures and triangulation across perspectives). NIME+1 Brown et al.'s meta-review of music interaction evaluations further supports your motivation by documenting how much evaluation remains informal and which UX dimensions are typically under-examined. teamaxe.co.uk Your contribution to this line is the modality-access manipulation (visual-only/audio-only/audiovisual) as a direct operationalisation of intermedial equality and interference, plus the linkage of participant experience to independent clip ratings—an approach that is legible to both Organised Sound and NIME/SMC audiences. 2.7. Tonnetz and cellular automata as compositional substrates The Tonnetz has a substantial theoretical lineage in neo-Riemannian and transformational theory, and Cohn's formalisation of Tonnetz spaces and parsimonious operations is a canonical reference for grounding your mapping as more than a "clever layout". repmus.ircam.fr+1 In your system, the Tonnetz is not simply a visualisation of harmonic relations; it is an interaction scaffold that makes local neighbourhood operations musically meaningful under constrained exploration. Cellular automata have likewise been explored as generative musical formalisms for decades. Miranda's early work on music composition using cellular automata is a clear historical anchor

for treating CA as a legitimate compositional engine rather than a novelty. Plymouth Research Portal Earlier ICMC work also demonstrates the long-standing interest in CA as a bridge between simple rules and emergent musical structure. Quod This tradition provides the backdrop against which your system's distinctive move becomes clear: you are not merely using CA to generate notes; you are using CA as an intermedial score whose state is rendered simultaneously as sound and light, enabling interference to be studied as an emergent property of a coupled system.

3. Related work 2.1. Intermediality as a relational claim (not "multimedia" as co-presence) Intermediality is frequently invoked to describe work that happens "between" media, but its analytical value depends on how precisely the relevant intermedial phenomenon is specified. Rajewsky argues that intermediality functions best as a lens for describing boundary-crossing phenomena, and she proposes a typology—medial transposition, media combination, and intermedial reference—to avoid collapsing distinct practices into a single umbrella term. Érudit+1 In this paper, the study is primarily concerned with media combination (sound and time-based visual output co-present) and with the ways intermedial meaning is produced when neither medium is treated as explanatory "support" for the other. The Organised Sound thematic call intensifies this point by framing intermediality as a shift from juxtaposition to fusion, and by explicitly positioning sound and image as equals within an "Intermedial Space" in which meaning is mutually shaped rather than hierarchically assigned. Cambridge University Press & Assessment This paper takes that equality claim as the central problem to be investigated: not whether sound and image can be presented together, but whether they can be composed and perceived as co-constitutive. 2.2. "Media equality" across modalities: Elleström's model as an analytic scaffold To discuss equality without treating media as undifferentiated streams, Elleström's modality model is useful because it specifies how media relate through material, sensorial, spatiotemporal, and semiotic modalities. Springer Nature Link This matters for intermedial equality because equality can fail (or succeed) at different levels. A work may be equal in material investment (both present), but unequal in sensorial salience (one dominates attention), or equal in salience but unequal in semiotic authority (one is treated as the "real" carrier of form). In the present system, sound and light are produced as parallel projections of a shared evolving generative substrate. This establishes a form of causal parity (both modalities originate from the same state-space), but does not guarantee perceptual or interpretive parity. Elleström's framework therefore supports the call's demand for reconceptualisation by allowing the paper to ask: in what modalities is equality achieved, and where does it collapse into hierarchy? Springer Nature Link+1 2.3. Intermedial Space and intermedial interference The call foregrounds Intermedial Interference—the constructive and/or destructive interaction of different media features—as a prompt for new critical frame-

works. Cambridge University Press & Assessment Chiaramonte's doctoral research develops "intermedial interference" in the specific context of electroacoustic audiovisual composition and positions the intermedial as a compositional site: interference describes how interactions between media features (e.g., temporal articulation, density, motion, segmentation cues) can amplify meaning or destabilise it. Bournemouth University Research Online+1 For this paper, intermedial interference is treated as an emergent property of a coupled system, and therefore as design-sensitive: mapping strategies, constraint regimes, and interaction affordances can shift whether interference is experienced as productive fusion or as competition/contradiction. This framing motivates an evaluation approach that does not reduce intermediality to synchronisation accuracy, but instead examines how participants shape (and retrospectively interpret) the fused audiovisual artefact under different access conditions. 2.4. Tonnetz "misleadingness" as productive interference: cognitive/HCI grounding A further complication—especially relevant to Tonnetz-based instruments—is that the visual organisation is not a neutral display of theory; it can actively shape users' mental models and expectations. Besada et al. examine the Tonnetz "at first sight" as a pitch-space interface through a cognitive/HCI lens, showing that while theoretical competence supports partial comprehension of Tonnetz structure, the geometry can also bias understanding, particularly when sequences do not align with functional-harmonic expectations. SAGE Journals+1 This result is directly valuable to an intermedial interference argument. In a system where the Tonnetz is simultaneously a harmonic scaffold and a time-based visual field, users may "overfit" the geometry—reading theoretical meaning into visually salient patterns. Such overfitting can be framed as destructive interference when it produces misleading causality or contradictory form cues, but it can equally be framed as constructive interference when it creates a generative space for exploratory interpretation and novel compositional direction. The present study is designed to operationalise this tension: if Tonnetz-driven visual structure shapes musical expectation, then composing with one modality withheld (audio-only or visual-only) and then replaying the full audiovisual result becomes a controlled way to observe how expectations are formed, violated, reconciled, and valued. 2.5. Remediation and the oscillation between transparency and foregrounded mediation Rajewsky explicitly compares intermediality's typology to Bolter and Grusin's account of remediation, which describes how new media refashion prior media and how experiences can oscillate between transparency (immediacy) and conspicuous mediation (hypermediacy). Érudit+1 For generative audiovisual instruments, this oscillation is not merely historical; it is experiential. Interaction may feel transparent when cause-effect relations are graspable ("I am shaping the piece"), and hypermediated when the system's autonomy becomes perceptually foregrounded ("the system is doing things to me"). This shift is a mechanism through which media hierarchy and interference are produced in practice, and it reinforces the

need for an evaluation approach that captures both felt agency and perceived coherence of the resulting artefacts. 2.6. Positioning the present study Taken together, these intermedial frameworks (Rajewsky; Elleström; the Organised Sound call; Chiaramonte) motivate treating "media equality" and "intermedial interference" as empirically tractable, without collapsing them into simple preference judgements. Bournemouth University Research Online+3Érudit+3Springer Nature Link+3 The study therefore evaluates how intermedial meaning-making behaves under controlled modality access, and how Tonnetz-driven visual structure may function as both a biasing force and a creative resource (Besada et al.). SAGE Journals+1

4. Method 3.1. Participants Sampling and recruitment. Participants were recruited via the author's professional and personal networks using convenience sampling. Recruitment sought a mix of musical and non-musical backgrounds, and a range of familiarity with generative systems. Sample. N = [ ] participants completed the solo protocol. Participants self-reported: age range; musical experience (none / some / moderate / high); music theory / harmony familiarity; familiarity with generative or algorithmic music systems; Tonnetz familiarity; perceptual notes (colour-vision deficiency, sensitivity to flashing lights, optional comments). Dyad subset. N_dyad = [ ] dyads completed an exploratory collaboration condition (Section 3.3.3). Ethics and consent. All participants provided informed consent for audio/video recording and for anonymised excerpts to be used in a subsequent external rating task.

3.2. System and apparatus 3.2.1. Instrument under study The study evaluates the TZ5 / Cellular Au-Tonnetz audiovisual instrument: a unified system in which sound and light are generated simultaneously from a shared Tonnetz–cellular automata (CA) substrate. The system has been previously documented as a technology description (rather than an evaluation), providing a stable technical baseline for the present study. The TZ5 system comprises four main components: Web application (UI + configuration store): a browser-based interface that captures user inputs (sliders, buttons) and stores parameter values in an SQL database; the latest configuration is served to the device via HTTP GET as JSON. Control board: an ESP32-S3 microcontroller responsible for CA state updates and for driving both musical and visual outputs based on the retrieved parameters. Audio synthesis subsystem: renders audio from the device's MIDI output, either via a DAW (e.g., Ableton Live) or via a hardware synthesiser. Hexagonal LED panel subsystem: a 91-pixel addressable LED array (single serial data line protocol such as WS2811) displaying the evolving CA state. [Figure 1 about here: system overview / block diagram showing web app → ESP32-S3 controller → LED panel + MIDI → Ableton.] [Figure 2 about here: web UI screenshot showing the parameter set used in the study.] 3.2.2. Coupled audiovisual generation (shared state-space) Sound and light are produced as parallel projections of the same evolving CA/Tonnetz state. The controller assigns MIDI note numbers to Tonnetz grid cells (virtual "orches-

tra members"), then evaluates CA rules and population limits to determine activations; cells that become active both (i) trigger MIDI note events and (ii) update corresponding LED pixels. Note–colour mapping. Each pitch class is mapped to a hue using an HSV colour model; the LED controller updates RGB values as a visual representation of harmonic activity. 3.2.3. LED panel configuration The control board can drive different physical panel formats provided the LED count and protocol remain consistent. The study panel was: Panel type: [91-LED custom PCB panel (10 cm) / 70 cm plywood panel with 91 pixels] Diffusion: [paper / opaque acrylic / other] (held constant across sessions) [Figure 3 about here: LED panel photograph (front view).] [Figure 4 about here: LED panel in situ / participant viewpoint (optional).] 3.2.4. Participant-facing control parameters Participants controlled the instrument via the web UI, adjusting parameters documented as core TZ5 features. Parameter labels below follow the UI conventions used in the system documentation. Population and density Max Notes (Max Population): caps the maximum number of active orchestra members (density limit). Min Notes (Min Population): enforces a minimum active population; if activity collapses below this threshold the system seeds new activations to maintain baseline activity. Temporal dynamics Rate: CA update rate (selectable in milliseconds; real-time adjustable). Life Length: maximum lifetime of an activation (implementation-specific; reported as configured in the firmware revision used). Neighbourhood and rule constraints Neighbour Counting Method (Local / Extended): Local counts 6 neighbours; Extended counts 18 neighbours. Min Neighbours: minimum neighbour threshold required for activation. Max Neighbours: maximum neighbour threshold preventing activation when exceeded (crowding constraint). Recurrence and memory Loop On/Off: toggles between random coordinate selection and replay of stored coordinate selections. Loop Steps (Loop Length): number of updates stored and replayed; loop duration is determined by the update rate and loop steps. Harmonic admissibility Scale selection: constrains which pitch classes can activate; the same constraints shape the corresponding colour palette because colours are mapped to pitch classes. Excluded to reduce confounds Chord progressions / automatic scale sequencing were disabled for this study, so that formal change is attributable to participant steering rather than automation. (Chord progressions are a documented feature of TZ5 but were not used here.) 3.2.5. Audio/video capture and synchronisation All conditions produce paired audiovisual artefacts, regardless of which modality the participant could access during composition. Audio capture: [Ableton Live direct render / line-out capture], [sample rate/bit depth]. (Ableton is explicitly supported as a rendering target for TZ5's MIDI output.) Video capture: fixed camera framing the panel at [resolution/frame rate], with fixed exposure and white balance. Synchronisation marker: a clap and/or UI "MARK" at the start of each captured excerpt. Traceability: each excerpt is associated with a participant ID, condition ID, saved preset ID, and firmware revision ID. [Figure 5 about here: recording setup photograph/schematic (panel, camera, participant position).] 3.2.6. Replication artefacts and supplementary materials To support reproducibility and independent interrogation of system behaviour, the following

artefacts will be provided as supplementary materials: Supplementary S1 (Ableton Live project): Ableton Live project containing the synthesis patch, MIDI routing, and recording configuration used for all sessions. Supplementary S2 (Reference audio renders): stereo reference renders of the baseline preset (S0) and a short calibration sequence for level/routing verification across systems. Supplementary S3 (Parameter presets): exported parameter sets (S0 and all end-state presets) in a machine-readable format (e.g., JSON), keyed to participant ID and condition. (The web application architecture explicitly serves/stores parameters as JSON retrieved by the device.) Supplementary S4 (Firmware + build/flash instructions): the microcontroller firmware revision deployed in the study, including build/flash instructions and a version identifier corresponding to the device. This enables verification of CA update logic (asynchronous update strategy), neighbour counting (local vs extended), looping behaviour, and note–colour mapping. Supplementary S5 (Stimulus set): anonymised audiovisual clips supplied to external raters (including baseline control clips), with clip IDs corresponding to the analysis dataset.

3.3. Study design 3.3.1. Overview The study comprises: a within-subject solo protocol with three task/condition blocks manipulating modality access during composition; an exploratory dyad protocol with asymmetric modality access; and a system-alone baseline capture recorded from the shared starting preset. Dyad-only sessions skipped the solo blocks and end-of-session comparison, proceeding directly to the dyad questionnaire and final reflections. 3.3.2. Baseline control (system-alone) For each session, the system ran from the common starting preset (S0) for the same duration as a participant block with no parameter changes. A final excerpt was captured, producing a baseline artefact representing system behaviour absent human steering. 3.3.3. Dyad condition (exploratory): asymmetric modality access collaboration Dyads were seated facing each other to support negotiation. One participant had auditory access and the other had visual access, while both shared access to the parameter interface. Audio-role participant: wore headphones; did not see the LED panel. Visual-role participant: faced the LED panel; had no access to audio. The dyad produced a single final excerpt from S0 after collaborative parameter steering.

3.4. Procedure 3.4.1. Orientation and familiarisation Participants received a briefing describing the instrument as a coupled audiovisual generator and explaining the three task blocks. At the start of the session, a single familiarisation phase (approximately 10–15 minutes) with full audiovisual access allowed participants to develop a working understanding of key parameters (density, update rate, neighbourhood method, looping, scale constraints). Session metadata recorded the participant ID and the counterbalanced block order (A/B/C), with an option to flag dyad-only sessions. 3.4.2. Common starting point (S0) Each block began from a standardised preset S0 to enable within-subject comparison. The system supports user-defined initial states; in this study S0 used [central pixel active / specified initial state]. 3.4.3. Solo blocks (three tasks with modality constraints) Each participant completed three blocks. In each block the participant explored the system for approximately 5–10 minutes under the

specified modality constraint, then recorded a final excerpt of [60–120] seconds. Participants were permitted to adjust parameters during the recorded excerpt. The end-state preset was saved. Block A: Visual-only composition Constraint: participant viewed the LED panel; audio feedback was fully masked. Task prompt: "Create an interesting pattern/motion on the LED panel." Output: final captured excerpt includes both audio and video (audio revealed only after capture). Block B: Audio-only composition Constraint: participant heard audio output via headphones; the LED panel was occluded. Task prompt: "Create an interesting melody/sonic texture." Output: final captured excerpt includes both audio and video (video revealed only after capture). Block C: Audiovisual composition Constraint: participant had simultaneous access to audio and the LED panel. Task prompt: "Create an interesting combined audiovisual experience." Output: final captured excerpt includes both audio and video. Order control. Condition order was counterbalanced across participants using a rotating schedule. Reveal + questionnaires (per block). Immediately after each capture, participants completed a pre-reveal questionnaire (Part A). The excerpt was then replayed with both modalities enabled and participants completed a post-reveal questionnaire (Part B), including free-text reflection comparing expectations to outcome and notes on intermedial interference. 3.4.4. Dyad trial (single collaborative block) Dyads completed one collaborative block from S0: confirm asymmetric access and shared UI control; exploration under split access (~5–10 minutes); final capture ([60–120] seconds) with live parameter changes permitted; save end-state preset; complete dyad questionnaire and brief joint reflection.

3.5. Measures 3.5.1. Block questionnaires (pre-reveal and post-reveal) Each solo block used a two-part questionnaire. Part A (pre-reveal) captured task focus (stability/clarity, complexity/energy, or mixed), strategy notes, and 7-point Likert ratings of satisfaction, intention clarity, steerability, interface understanding, useful surprise, frustrating unpredictability, and anticipated interest. Part B (post-reveal) captured 7-point Likert ratings of fusion/equality, coherence, constructive and destructive interference, overload, expectation match, interpretation change after reveal, causal legibility, perceived system autonomy, and reliance on visual or Tonnetz/theory cues. Participants also provided free-text reflections on expectation vs outcome, noted any interference moments, and selected the most influential parameters (aiming for three). 3.5.2. End-of-session comparison Participants ranked their three outputs and indicated which block felt most intermedial, which block had the largest expectation–outcome mismatch, and provided a short reflection on what they learned about sound–light relationships, plus one suggested change to improve media equality. 3.5.3. Dyad questionnaire (if applicable) Dyad participants provided ratings on communication effectiveness, shared reference points, whether split access helped or hindered coordination, sense of joint ownership, audiovisual balance/coherence, and whether the dyad outcome would be preferred over solo outputs. They also recorded dyad ID, role (audio-role or visual-role), dyad preset ID, and brief notes on communication and disagreements. 3.5.4. Final reflections (optional adden-

dum) An optional end-of-session addendum captured a title and one-line description for the participant's favourite piece, authorship attribution and rationale, return likelihood and conditions, imagined contexts of use, target user profile, suggested removals/additions, collaboration expectations (easier to negotiate visuals vs notes), and confidence in recreating a similar result. 3.5.5. External rating study (independent raters) A separate pool of raters evaluated anonymised audiovisual clips in a blinded presentation: Stimuli: one clip per participant per block, plus system-alone or default-parameter control clips. Presentation: a tokenised web app generated a randomised clip order; raters completed an information/consent page before access and were instructed to allow 1–2 hours to rate approximately 30 clips (1–2 minutes each). Ratings were unlocked only after each clip was watched to the end. Judgements: 7-point Likert ratings of preference, coherence, novelty, fusion/equality, constructive/destructive interference, overload, inferred structure/process, and perceived human steering; separate ratings of memorability and perceived agency; a best-context choice (installation, live performance, background ambience, workshop/education, unsure); attention-dominance selection (sound vs light vs balanced); a composition-condition guess (audio-only, visual-only, audiovisual, unsure); and optional comments. An optional end-of-session form captured raters' top three clips and brief fusion notes.

3.6. Data handling and analysis plan Per block data captured saved preset (parameter values); audio file + video file; questionnaire responses; [optional] addendum reflections and researcher field notes; [optional] parameter change logs from the server database (where available). Primary comparisons within-subject differences across visual-only vs audio-only vs audiovisual blocks; participant-shaped outputs vs system-alone baseline; exploratory dyad outputs vs solo outputs (reported descriptively with appropriate caution given sample size). Cross-linking experience to artefacts Participant self-ratings (e.g., fusion, agency, interference) will be compared with independent rater judgements of the corresponding clips to test whether perceived intermedial success is primarily a compositional experience, an artefact property, or both.

4. Results: composing and interpreting intermedial relations under modality constraint 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 across participants. 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. Table 1 summarises participant background metadata.

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 inter-

medial" (7/9), with B selected by 1/9 and none selecting A (one response was recorded as unsure depending on coding). 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 Fig1_outcomes_counts.

This pattern is consistent with a broader interpretation evident in the qualitative accounts: perceived intermediality is not simply a judgement of "mapping accuracy." Rather, it is an experiential judgement about whether the coupled system supports (i) intention formation, (ii) monitoring and causal inference during action, and (iii) retrospective sense-making during replay.

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

Across all conditions, participants described strategies that clustered into five recurring forms:

Density shaping, for example "filling the grid then thinning it" and returning to higher density near the end;

Arc/narrative form, typically described as build-up toward chaos followed by a release or resolution;

Loop-based structuring, using loop length and loop on/off as a "container" to stabilise material and manage risk;

Harmony/scale manipulation, sometimes via tonal intention and sometimes via colour similarity as a proxy for consonance;

Exploratory 'A/B testing', rapidly toggling parameters or notes to learn local cause–effect relations.

However, the relative emphasis of these strategies shifted systematically with modality constraint:

Visual-only (A) encouraged composition through spatial and density cues. Participants often treated the panel as a dynamic visual field, composing by "covering the grid," sculpting density, and exploring emergent pattern families. Consistent with this, perceived steerability (A3) was highest in A (median 6.00 [4.00, 6.00]; Table 2; Fig2_paired_control_A3), suggesting that visual feedback supported a more confident action–evaluation loop.

Audio-only (B) more often induced strategy framed in texture, pacing, and risk management. Participants described looping as a means of establishing predictability and preventing collapse into uncontrolled noise. Quantitatively, steerability dropped in B (A3: 4.00 [3.00, 5.00]), aligning with repeated descriptions of reduced state visibility and uncertainty about what the system was doing.

Audiovisual (C) functioned as a synthesis condition: participants commonly reported combining what they learned in the earlier blocks and adopting a more

deliberate alternation between density control, harmonic selection, and temporal shaping. Satisfaction (A1) was highest in C (6.00 [5.00, 6.00]) and useful surprise (A5) also increased (6.00 [5.00, 6.00]; Table 2), supporting the qualitative framing of C as enabling both intention and productive emergence. At the same time, accounts also note the potential workload of trying to optimise sound and light simultaneously; frustration from 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), consistent with a "two-objective" monitoring challenge.

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

A recurring phenomenon across participants was the reveal (replay with both modalities enabled after composing under constraint) acting as a moment of retrospective binding and reinterpretation. Participants frequently described surprise and delight on replay, and reported that the coupled output could feel more coherent than the compositional experience itself. Several accounts explicitly point to replay as disclosing latent structure in the shared generative substrate (e.g., unexpected pattern regularities, clearer musical phrasing than felt "in the moment").

This is supported quantitatively by the intermediality 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), and summarised in Fig3_paired_intermediality_index. In other words, participants were most likely to judge the reveal as "the same process" when sound had been salient during composition (B) or when both modalities were available (C), whereas composing visually tended to produce weaker same-process judgements on reveal for some participants.

Crucially, these results clarify why the audio-only block was often selected as the biggest mismatch (6/9; Fig 1) without necessarily being evaluated negatively. In multiple accounts, mismatch operated as a productive discontinuity: replay with both modalities could exceed expectations, functioning as a reward and 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 their experience and that audio had to be removed to fully prioritise visual composition. Others self-identified as visually oriented and described greater controllability when composing visually, while finding audio-only harder to control despite later recognising more structure on replay.

Quantitatively, "balanced modalities" (B2) had 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 matches the qualitative pattern that "media equality" is not an intrinsic property of the instrument; it is negotiated through (i) attentional dominance, (ii) familiarity, and (iii) the availability of stabilising cues (especially looped repetition).

Notably, 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). This is important because participants sometimes used terms like "chaos" or "busy," but the quantitative ratings suggest that overload was not the dominant experiential failure mode at the group level. Instead, the more diagnostic experiential issue was predictability and monitoring (mismatch, cue conflict, and interpretability), particularly in B.

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, in which one modality implied temporal or causal relationships that the other did not support.

The most concrete and recurrent cue conflict concerned duration: participants noted that light persistence could be misleading relative to note length, undermining temporal correspondence and weakening cross-modal prediction. A second cue conflict concerned density: sparse visuals were sometimes perceived as sonically "busy," while visually dense states could be less complex sonically than expected. These density-based mismatches are consistent with the system's generative mechanics (where local rule cascades and octave stacking can decouple perceived "count" of events from apparent spatial activity), but the key point is interpretive: participants were often attempting to infer sonic complexity from visual density (and vice versa), and the mapping did not always support that inference reliably.

A third issue was perceptual masking introduced by the synthesis choice: pure tones and octave stacking were described as collapsing multiple events into a single percept, reducing the participant's ability to audit harmonic relationships and thereby weakening cross-modal legibility. These accounts connect directly to both the "biggest mismatch" concentration in B (Fig 1) and the lower steerability in B (A3; Fig 2): when withheld or weakened cues reduce event discriminability, participants' internal causal model becomes less stable during action.

4.7 Agency, autonomy, and a co-creative stance toward an autonomous substrate (Fig 2; Table 2)

Across conditions, participants framed interaction less as direct control and more as negotiated agency. Multiple accounts emphasised the pleasure of being able to steer the system while remaining open to surprise. One participant explicitly questioned authorship ("whether I performed or the Tonnetz performed"), capturing a recurring ambiguity: creative credit and perceived agency fluctuated

with the system's responsiveness to intention.

This theme is consistent with quantitative 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 (C) often supported a shift toward a "co-creative" stance—acceptance and collaboration with the instrument—particularly after participants experienced both constrained conditions and developed a more stable causal model. This stance was frequently associated with higher creative satisfaction and reduced monitoring for mismatch during replay.

4.8 Parameter strategy and interpretability: what participants used to steer (Fig 6; Table 2)

Participants' self-reported "most influential parameters" show both a stable core and a modality-dependent shift (Fig6_param_influence_by_condition). Rate was the dominant control across all conditions (A: 8; B: 7; C: 6 selections), aligning with widespread use of tempo/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), consistent with participants foregrounding harmonic selection when auditory evaluation was possible in real time.

Other parameters (Life Length; neighbourhood and population thresholds; loop length/on-off) were cited at moderate frequency across conditions, supporting the qualitative "constraint tuning" pattern in which participants searched for stable regimes and then explored local variations. In qualitative terms, this often manifested as discovering "stable states" separated by bursts of disorganisation, then using loops and rate to manage transitions between them.

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

Participants produced convergent, instrument-actionable requirements that map directly onto the mechanisms identified above:

Loop and reset affordances: clearer loop behaviour, predictable toggling, and reset/preset functions to reduce state uncertainty and support rapid recovery from unintended regimes.

Parameter semantics and scaling: units, non-linear scales where appropriate, and correction of counter-intuitive mappings (notably "rate").

Interface cohesion: reduced fragmentation (e.g., single-page layout without scrolling) and improved ergonomic/tactile interaction to support moment-to-moment correction.

Colour fidelity and mapping legibility: closer match between controller colours and panel colours; improved discriminability for near hues; optional persistent colour indicators for inactive notes to support harmonic planning.

Temporal correspondence: alignment between note duration and light persistence, or an intentional intensity decay to communicate note release (and to reduce duration-based cue conflict).

Richer articulation and event discriminability: ADSR/envelopes and/or richer timbres to reduce perceptual masking and improve the auditability of stacked events.

These requirements connect back to the results above in a straightforward way: where mapping legibility and temporal cues were ambiguous, participants reported mismatch and reduced controllability; where cues supported stable reference (especially looping), participants described safety, predictability, and more intentional structuring. In this sense, "intermedial interference" in this system is not solely a perceptual phenomenon; it is also shaped by the transparency of system controls and the reliability of the cues required for iterative prediction–action–evaluation.

Summary. Taken together, the merged results suggest that intermedial relations in this instrument are experienced as a dynamic negotiation between cue reliability, attentional dominance, and an evolving causal model of a shared generative substrate. Constraint conditions do not simply degrade the experience; they reconfigure how participants form intention, monitor outcomes, and attribute agency—sometimes producing negative mismatch (loss of legibility), and sometimes producing positive mismatch (reveal-driven rebinding and delight).

5. Discussion
6. Conclusion
7. Acknowledgements

References

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. Springer Nature Link+1 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. SAGE Journals+1 Bolter, J. D. and Grusin, R. 1999. Remediation: Understanding New Media. Cambridge, MA: MIT Press. 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. Bournemouth University Research Online+1 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. Springer Nature Link 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). Cambridge University Press & Assessment Rajewsky, I. O. 2005. Intermédialité, intertextualité et remédiation: une perspective littéraire sur l'intermédialité. Intermédialités 6: 43–64. https://doi.org/10.7202/1005505ar. Érudit+1