Organised Sound (2026), 1–11
doi:10.1017/S1355771826101605

CAMBRIDGE UNIVERSITY PRESS

Check for updates

Article

Intermediality beyond synchronisation: Text as a compositional agent in electroacoustic audiovisual performance

Patrick Hartono
School of Communication and Design, RMIT University, Ho Chi Minh City, Vietnam

Abstract

This paper examines how text can operate as an intermedial compositional agent in electroacoustic audiovisual performance. Through a focused practice-based case study of a single performance – F2 – in which electroacoustic sound, procedural visuals and projected and spoken text unfold without synchronisation or audio-reactive coupling between the sonic and visual layers, the study considers text not as a thematic focus but as a structuring element that shapes perceptual meaning-making within this configuration. Using Michel Chion's concept of added value, the article argues that textual narration can amplify and reframe the perception of sonic and visual materials without relying on synchronisation or audio-reactive correspondence. Qualitative audience responses indicate that text functions as an interpretive anchor, guiding attention, emotional orientation and coherence within an otherwise unsynchronised audiovisual environment. On this basis, the paper proposes a triangular added-value model in which text functions as the primary directional mediator; text semantically anchors sonic intensity, metaphorically reframes visual abstraction and enables sound and image to be perceived as jointly significant despite their structural independence. The study contributes a practice-based perspective on how text can be integrated into electroacoustic audiovisual composition to expand approaches to form, perception and artistic communication in contemporary new media performance.

Keywords: electroacoustic audiovisual composition; intermediality; perceptual meaning-making; text as compositional agent; unsynchronised sound–image relations

(Received 15 January 2026 UTC; revised 20 August 2026 UTC; accepted 1 September 2026 UTC)

1. Introduction

1.1. Research focus

The focus of this research is to examine, through a focused practice-based case study, how text can operate as a primary structuring force within an electroacoustic audiovisual composition where sound and image are intentionally left without synchronisation. Rather than relying on parameter-mapping or audio-reactive strategies that have shaped much recent audiovisual practice, this study explores how projected text within the context of live electronic performance can guide audience perception and shape the expressive interpretation of electroacoustic sound and procedural visual materials. The central question concerns how text influences meaning-making in this specific configuration, where no technical or temporal alignment is used to connect the sonic and visual domains. In contrast to a substantial body of work that theorises electroacoustic audiovisual practice primarily through synchronisation and mapping paradigms, the role of text as a primary structuring layer in such unsynchronised configurations remains comparatively under-theorised. Although this article does not undertake a comparative analysis of synchronised and unsynchronised contexts, it addresses this under-theorisation by examining one such configuration in depth, foregrounding textual narration as a compositional agent.

By analysing a performance where textual cues orient attention, frame emotional intensity and support coherence in an otherwise unsynchronised audiovisual environment, the research aims to demonstrate how the presence of text can help the work achieve Michel Chion's concept of added value, not through synchronisation or audiovisual correspondence, but through perceptual anchoring and interpretative guidance. Through a combination of practice-based research and audience response analysis, this study positions text as an active compositional agent within contemporary electroacoustic audiovisual performance and proposes a triangular understanding of meaning-making between sound, image and text.

This investigation also emerges from my long-term practice in audiovisual performance since 2011, during which I have explored varied approaches to sound–image interaction – including synchronised, reactive and generative systems – leading to the present shift towards an unsynchronised model in which text forms a parallel compositional layer.

1.2. Context within electroacoustic audiovisual practice

I treat algorithmic as the superordinate term, with procedural and generative as subcategories distinguished by the degree to which

Email: patrick.hartono@rmit.edu.vn

Cite this article: Hartono P. (2026). Intermediality beyond synchronisation: Text as a compositional agent in electroacoustic audiovisual performance. *Organised Sound*, 1–11, https://doi.org/10.1017/S1355771826101605

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

outcomes are specified in advance. Electroacoustic audiovisual composition increasingly incorporates real-time interactions between sound and moving image, often through algorithmic visual systems. Recent practice-based research highlights approaches in which electroacoustic sound informs or influences visual behaviour through parameter mapping, real-time analysis and procedural animation (Tsang 2023; Chiaramonte 2024). Within this broader field, audiovisual works are frequently developed in technological environments that encourage close structural alignment between sound and image, making synchronisation a common strategy for achieving perceptual cohesion.

Current creative software ecosystems further reinforce this orientation. Tools such as TouchDesigner, Ableton Live, Max for Live and Resolume provide integrated pathways for audio-reactive workflows that are routinely presented as primary methods for constructing audiovisual performance. Ableton's documentation, for instance, introduces TouchDesigner integration through tutorials focused on 'making audio-reactive visuals with Live', where real-time audio signals are mapped directly onto visual parameters. Similarly, the TouchDesigner platform disseminates workshops, educational programmes and community tutorials that foreground audio-reactive networks – often using Fast Fourier Transform (FFT) or amplitude data to modulate shaders and compositing pipelines – and present such setups as straightforward entry points for building audiovisual works. Collectively, these tools and learning resources contribute to a landscape in which synchronisation, typically achieved through audio analysis and parameter mapping, remains a dominant approach.

Correspondingly, recent literature in electroacoustic audiovisual practice often centres on sound–visual mapping strategies, visual–musical correspondences and systems designed to maintain tight temporal and structural relationships between modalities. In contrast, there is relatively limited discussion of text functioning as a core compositional element within electroacoustic audiovisual work, particularly in settings where sound and image are deliberately unsynchronised and text plays a perceptual or interpretative role rather than serving as subtitle, caption or linguistic annotation. In this paper, I aim to address this gap by examining how textual narration can operate as an active layer within electroacoustic audiovisual composition, supporting perceptual orientation and meaning-making without relying on synchronisation or audio-reactive methods. Against this backdrop, I offer a counter-example in the form of a practice that deliberately avoids structural synchronisation, instead using text as a binding element that mediates and organises the relationships between electroacoustic sound and procedural visuals.

1.3. Case study and scope

In this article, I draw on my recent electroacoustic audiovisual piece, 'F2'. The work integrates electroacoustic sound, procedural visual material and a parallel textual component delivered through both projected text and live spoken narration (vocal improvisation).

Importantly, no synchronisation or audio-reactive linkage is used between the sonic and visual layers; instead, each operates independently when the text functions as a separate compositional stream. I therefore do not examine the performance as an artwork in its own right, but as a research object that enables focused inquiry into how textual elements influence perceptual interpretation in an unsynchronised audiovisual environment.

The narrative content referenced within the text – drawn from a historical event in Indonesia involving state violence and the death of a child as a result of sexual violence – is not subjected to detailed historical or sociopolitical analysis here. Rather than treating this material as the primary thematic object of critique, I focus on its structural and perceptual functions: how it operates as compositional material that provides interpretative cues for the audience and contributes to the work's affective and narrative coherence. At the same time, I acknowledge that this content carries significant political and ethical weight, particularly in relation to trauma and revisionist histories, and that such weight forms an integral content layer that conditions how the textual narration is received. The scope of the discussion thus centres on electroacoustic composition, audiovisual practice and new-media performance methodologies, while recognising the ethical implications of staging such material in performance.

Methodologically, the study combines practice-based reflection with qualitative audience feedback collected after the performance. The aim is not to produce a sociological account of audience reception, but to understand how textual cues shape perceptual orientation, coherence and meaning-making when sound and image are intentionally unsynchronised. Within this qualitative audience study, I treat responses as a way to trace how listeners and viewers organise their experience around the textual narration, rather than as data for broader generalisation about audiences. This delimited scope ensures that the analysis remains focused on the role of text within electroacoustic audiovisual composition and its capacity to enable the work to achieve added value.

A note on scope is necessary here. This study does not undertake a comparative investigation of synchronised versus unsynchronised audiovisual configurations, nor does it claim that text functions in fundamentally different ways across these configurations. Rather, it offers a focused practice-based account of one configuration – F2 – in which sound, image and text are deliberately decoupled, and examines how text functions within this specific arrangement. The contribution lies in articulating the perceptual and compositional roles that text takes on under these conditions; the triangular added-value model developed in Section 5 should be read as informed by, rather than generalised from, this single case.

2. Background and conceptual framework

2.1. Electroacoustic audiovisual composition

I understand electroacoustic audiovisual composition as a practice in which sound and moving image are treated as coexisting materials within a single compositional space, rather than as accompaniment or illustration in a hierarchical relationship. This perspective extends concerns familiar from electroacoustic music – such as timbre, spatialisation and acousmatic listening – into multimodal contexts where visual form, motion and temporality become part of the same organisational field (Emmerson 2007). Within this broader lineage, audiovisual works are often approached as integrated environments in which sonic and visual elements are composed together through studio-based processes, live electronics or hybrid methods.

Recent electroacoustic audiovisual practices increasingly employ real-time algorithmic systems, both procedural and generative, that enable dynamic interactions between audio analysis, synthesis and visual procedures. Within such environments, shared data sources and mapping strategies are commonly used to

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

structure relationships between sound and image in performance and installation contexts (Ikeshiro 2012; Carvalho 2025).

At the same time, it has been recognised that software tools and computer music environments embody idiomatic tendencies that shape aesthetic and compositional decisions. The design of digital music tools can privilege particular interaction patterns, making tightly coupled or audio-reactive audiovisual configurations more immediately accessible than approaches based on looser intermodal relationships (McPherson and Tahıroğlu 2020).

2.1.1. Practice context: Hutan Plastik and earlier works

My own practice in electroacoustic and audiovisual composition, developed since 2011, is situated within this broader field. Across earlier audiovisual works such as *Matrix Studies*, *Noise to Signal* and *Parakletos*, I frequently employed audio-reactive strategies to couple sound and image, while also experimenting with moments in which the visuals could stand more independently or even become temporarily dominant. In these pieces, sections of strong audio–visual mapping would sometimes be interrupted by passages in which the image no longer responded directly to the sound, yet the overall structure of each work still relied on some form of synchronisation at key points.

A more recent work, *Hutan Plastik* (2024), made this exploration more explicit by dividing the piece into four movements, each exploring a different degree of coupling between sound and image (Figure 1). The first, introductory movement is fully audio-reactive. In the second movement, which addresses the destruction of forests in Papua for palm oil plantations, I combine synthesised visuals derived from photographic material of the forest with a soundscape recording by Papuan composer Markus Rumbino; here, sound and image unfold as temporally independent layers without audio-reactive linkage. The third, climactic movement returns to a strongly audio-reactive configuration, whereas the fourth movement alternates between semi-audio-reactive and non-audio-reactive passages.

This work led me to ask whether a piece could sustain coherence when sound and image remain unsynchronised throughout, and what additional layer might support such coherence. The present study examines a subsequent performance in which sound and visual layers are deliberately left unsynchronised, with text introduced as a third compositional element, forming a triangular relationship within electroacoustic audiovisual composition.

2.2. Text in audiovisual works

Text has occupied a varied but often secondary position within audiovisual art, particularly in work outside film, documentary and television. In early visual-music and many electroacoustic traditions, language is largely absent or appears only as spoken voice, whereas written text rarely functions as an explicit compositional layer. With the rise of screen-based media, digital typography and networked computation, however, text has increasingly been treated not only as a vehicle for linguistic meaning but also as a time-based visual material that can participate in performance and moving-image composition (Bachfischer et al. 2006)

More recent studies analyse kinetic typography in title sequences, visual learning environments and documentary contexts, emphasising how temporal modulation, spatial movement, typeface choice and motion effects contribute to meaning-making and audience engagement in time-based media (Luthfiasari and Fadillah 2023; Zainuddin et al. 2025). Building on this, later work in motion graphics and design theory has examined how kinetic typography extends the expressive palette of graphic design into the temporal domain, arguing that the motion of text adds an additional layer of emphasis, rhythm and affect beyond static layout (Brownie 2015). These accounts locate typography within an expanded field of audiovisual practice in which textual form, reading tempo and visual rhythm are co-articulated.

Text has also been examined within digital performance and live coding, where code-based notation is written, modified and sometimes projected in real time as part of the performance interface. In this context, text functions simultaneously as instruction set, compositional material and, at times, a visible component of the stage image. The live coding literature emphasises that algorithmic notation – often realised through textual code – is a central medium through which performers articulate and reshape audiovisual systems, and that the visibility of code can mediate relations between human and machine agencies in performance (Blackwell et al. 2022; Villaseñor-Ramírez 2023; Xambó and Roma 2024). Here, text operates as an active structuring element in technologically mediated performance rather than a purely discursive layer.

Alongside these visually and code-oriented developments, electroacoustic and sound-based practices have long engaged with language, voice and text as sonic material. Traditions of text-sound composition, radiophonic work and acousmatic narrative – exemplified by the work of Trevor Wishart (1996) – foreground spoken voice, recorded speech and textual fragments as elements that can be cut, layered and transformed within electroacoustic texture, blurring distinctions between semantic content and timbral design. Such practices demonstrate how text may function simultaneously as vocal gesture, narrative reference and structural marker, positioning language not merely as a graphic or typographic entity but as a malleable component of electroacoustic practice operating across sonic and visual domains.

Across these domains, text emerges as more than a narrative vehicle, functioning to cue attention, frame interpretation and modulate relationships between sound and image. Within electroacoustic audiovisual composition, however, text remains under-theorised as a compositional element: research on audiovisual interaction continues to prioritise mapping, synchronisation and crossmodal perception, whereas text is more often discussed as dialogue, subtitle or graphic design.

2.3. Added value and its relevance to electroacoustic audiovisual composition

Michel Chion's notion of added value describes the perceptual enrichment that occurs when sound and image are combined, whereby each element acquires expressive and semantic qualities that it would not possess in isolation (Chion 1994). Added value foregrounds the ways in which audiovisual relationships shape interpretation, suggesting that meaning arises not only from the intrinsic properties of sound and image but also from their conjunction within time-based media. Closely related is Chion's concept of synchresis – the spontaneous synthesis of sound and image that occurs when they are presented in temporal alignment – which emphasises the perceptual immediacy with which audiences fuse auditory and visual events into coherent units (Chion 1994). In film theory, these ideas have explained how sound stabilises or destabilises visual meaning, creates continuity across edits and guides attention within complex scenes.

Within electroacoustic audiovisual performance and installation, these concepts have been taken up in ways that extend beyond

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

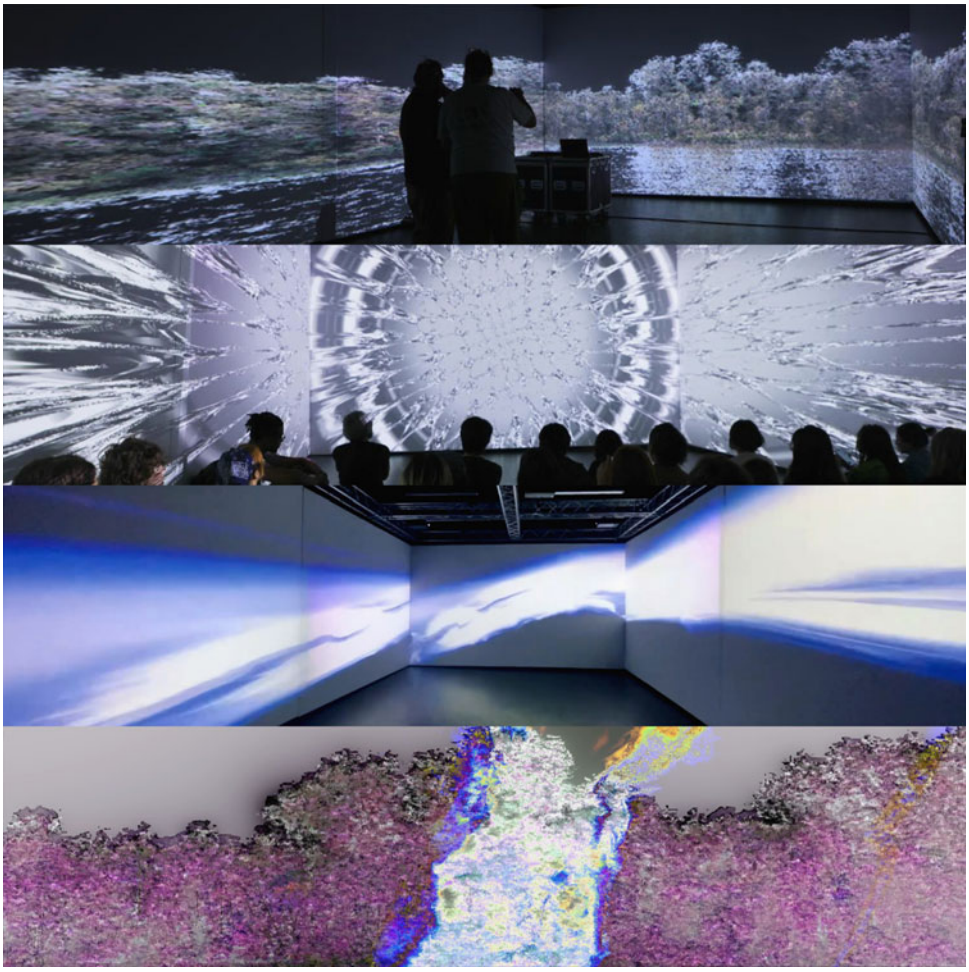


Figure 1. Hutan Plastik (2024): Four Movements. The works consist of movements that examine varying degrees of coupling between sound and image.

traditional cinema. Artists and researchers have examined how electroacoustic sound interacts with projection, light and spatialised visual forms to produce emergent perceptual wholes, often in environments that foreground materiality, spatial presence and embodied listening (Hill 2013; Basanta 2015; Zareei 2020; Chiaramonte 2024). In such contexts, added value can arise from the interplay between abstract sound fields and non-representational imagery, where the absence of explicit narrative invites audiences to construct their own interpretative frameworks. This has led to a range of intermedial practices in which sound and image are treated as co-constitutive forces, generating hybrid experiences that are neither purely musical nor purely visual.

At the same time, these works often rely – implicitly or explicitly – on some form of synchronisation or structural correspondence to support the emergence of audiovisual coherence. Even in settings where narrative is minimal or absent, synchresis remains a powerful mechanism; temporal alignment between sonic and visual events can create a sense of causality or gesture, encouraging audiences to perceive the two modalities as facets of a single phenomenon (Hill 2013; Basanta 2015; Zareei 2020). Yet, as recent practice demonstrates, coherence can also be sought through looser configurations in which sound and image do not share strict temporal coupling, but instead operate as parallel processes that interact perceptually through association, contrast or analogy (Chiaramonte 2024). These developments invite reconsideration of how added value might function in situations where synchresis is deliberately weakened or withheld.

In the context of this article, added value is understood not only as a property of sound–image relations but as a broader framework for thinking about how different media layers contribute to perceptual and interpretative richness in performance. Across earlier works, including *Hutan Plastik* (2024), I have experimented with ways in which electroacoustic sound and visual material can generate added value without relying solely on stable audio-reactive coupling. The case study examined here deliberately avoids audiovisual synchronisation, allowing electroacoustic sound and procedural visuals to unfold independently while introducing text as an additional compositional component. This opens the possibility that added value may emerge through interactions among three elements – sound, image and text – rather than through sound–image coupling alone. The triangular relationship between sound, image and text can thus be understood as an intermedial configuration in which meaning emerges through perceptual interaction rather than through hierarchical media alignment or technical synchronisation.

2.4. Supporting perceptual concepts

The argument developed in this article relies not only on compositional and theoretical frameworks, but also on established insights from perceptual research. In particular, the way text interacts with sound and image in the performance can be better understood through work on multisensory integration, crossmodal correspondences, attention and narrative framing.

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

A complementary line of research concerns crossmodal correspondences, the systematic associations people experience between features in different modalities, such as matching high-pitched sounds with small, bright or elevated visual objects. Spence's tutorial review synthesises a large body of evidence showing that such correspondences can guide perception and expectations across sensory channels (Spence 2011), and subsequent work has emphasised that such associations shape how audiovisual events are grouped, attended to and interpreted in time-based experience (Spence and Parise 2012). Even when explicit pitch–shape mappings are not employed, this research supports a broader view in which stimuli from different modalities are actively related by the perceiver rather than simply juxtaposed.

Attention is a further factor relevant to how textual elements might structure experience in an unsynchronised audiovisual environment. The way attention is directed over time, together with the cues that draw focus to particular aspects of a complex scene, can shape how viewers organise what they perceive – an observation I return to in the audience analysis when discussing how textual milestones appeared to coincide with reported shifts in listening.

Narrative theory offers a further perspective on how text may function as an interpretative frame rather than a fully developed story. Ryan's work on narrative across media describes how narrative cues can be distributed across different forms – verbal, visual, musical and digital – without any single medium carrying the entirety of the narrative burden (Ryan 2004). From this standpoint, short textual passages or fragments can operate as narrative or conceptual prompts that shape how audiences organise and make sense of otherwise abstract materials. In the context of electroacoustic audiovisual composition, such textual cues can therefore provide orientation and coherence without transforming the work into a conventionally narrative piece.

Taken together, these perspectives provide a vocabulary for the perceptual readings discussed in the audience analysis. Multisensory integration and crossmodal correspondence describe how stimuli from different modalities can be related by perceivers, and attentional and narrative frameworks describe how sparse textual cues can orient attention and frame interpretation. In the case study examined here, I draw on these perspectives to read how projected and spoken text function within F2 as a perceptual organiser – directing attention, providing semantic anchors for abstract electroacoustic and procedural materials and supporting the emergence of coherence and added value across the work as a whole.

3. Artistic methodology and work description

3.1. Compositional design

In this work, I structure the performance around three autonomous components – an electroacoustic sound process, a procedural visual system and a projected text layer – which unfold simultaneously within the performance space over a continuous duration of approximately ten minutes. Although presented together, these components are not technically synchronised, and each maintains its own temporal logic.

3.1.1. Live sound synthesis system

For the electroacoustic layer, I use real-time processing of my improvised spoken voice through concatenative synthesis in SuperCollider (Figures 2 and 3). As I deliver the narration live, fragments of the voice are continuously captured and incorporated into a dynamic corpus, unless I intentionally freeze the buffer. Multiple synthesis configurations operate in parallel, generating textures that shift between granular vocal residues, noise-inflected gestures and layered timbral transformations. The resulting sound structure is shaped by a combination of live input, algorithmic behaviour and stochastic modulation, creating a sonic environment that evolves independently of the visual layers. My vocal delivery therefore functions simultaneously as narrative medium and as raw material for electroacoustic elaboration.

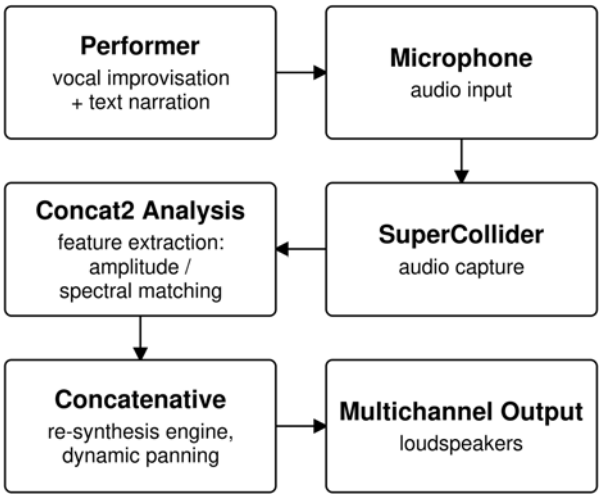


Figure 2. Interaction flow of the performance system: vocal input captured in SuperCollider, analysed via concatenative processes, resynthesised and routed to multichannel audio output.

3.1.2. Real-time procedural visual system

The procedural visual layer, which I project on the left side of the screen, is constructed from documentary footage of the May 1998 riots in Indonesia, transformed in TouchDesigner through a pixel-to-geometry conversion process that yields instancing-based, computer-generated patterns. This method produces abstract patterns and motion fields that, for me, retain the affective presence of the historical material while withholding recognisable imagery. I run the visual system continuously on a beats per minute (BPM)-based timeline without sending it any information from the audio or text. As a result, it develops through its own formal trajectory, generating motion, density and spatial rhythm independent of my performance and the sound processes.

On the right side of the projection surface, I present the projected text layer. It consists of a sequence of textual fragments that unfold autonomously according to a programmed timeline. In these texts, I address the historical background of the 1998 riots, current political discourse surrounding historical revision, and the story of Fransisca, an eleven-year-old victim of sexual violence during the riots as documented in contemporary reporting. I treat the projected text as a visual and semantic anchor within the performance, while keeping it structurally independent from both the procedural visuals and the electroacoustic sound (Figure 4).

The distribution of sound, image and text across separate yet co-present layers already implies an intermedial configuration in which relationships emerge through perception rather than through technical coupling. In this sense, the design of the work anticipates the triangular model of sound–image–text relations that I develop later in the analysis, where each medium functions as

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

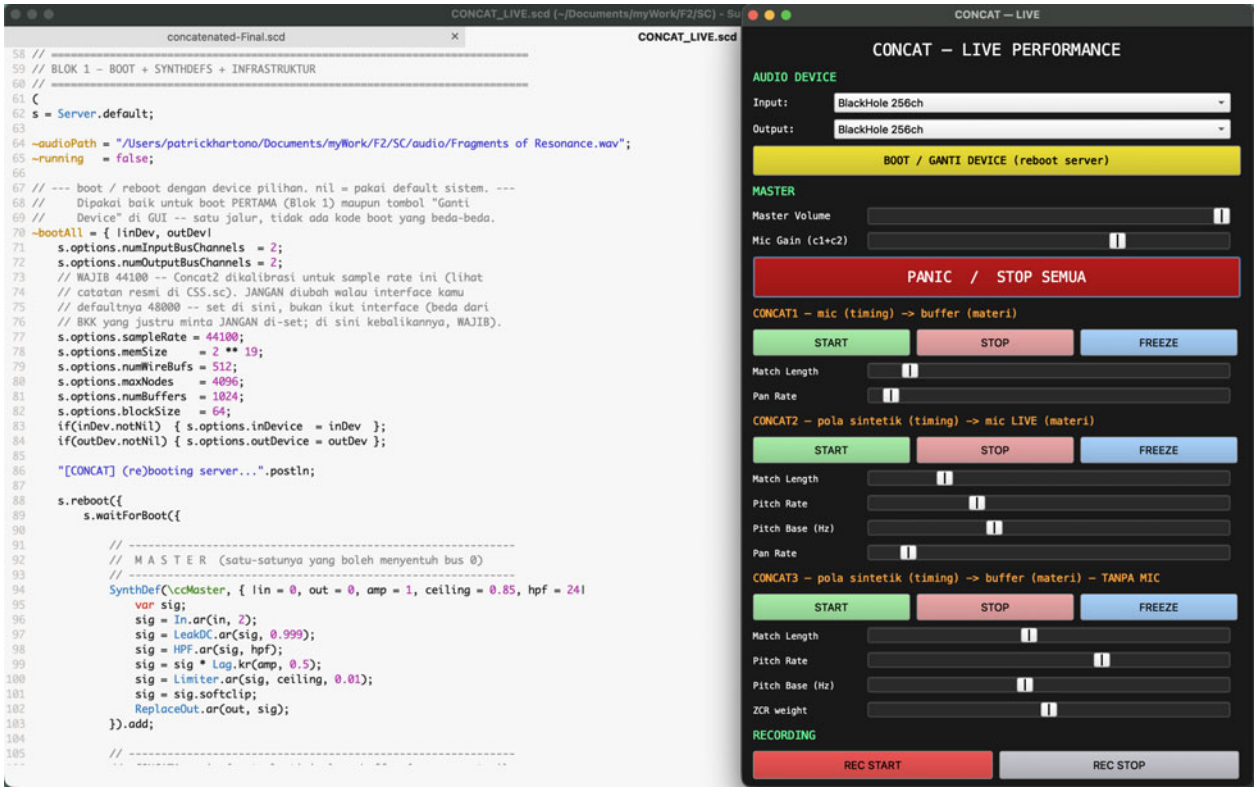


Figure 3. The live electroacoustic system: SuperCollider source (left) and the performance control interface (right), running three concatenative configurations in parallel. Interface labels are in Indonesian. Project files: https://github.com/patrickhartono/F2.

an independent stream that contributes to coherence and added value through their combined presence.

3.2. Non-synchronisation strategy

A central methodological decision in this work is my deliberate choice to keep audio, procedural visuals and projected text unsynchronised. I do not implement any mapping, cueing or shared parameter system across the three layers. Instead, each element develops according to its own temporal system: the electroacoustic sound responds solely to live vocal input; the procedural visuals follow their BPM-driven timeline; and the projected text unfolds according to a scripted progression.

This non-synchronisation approach functions both aesthetically and conceptually. The historical materials referenced in the work – particularly the traumatic events of 1998 and their subsequent political contestation – do not constitute a unified narrative. Testimonies conflict, documentation is incomplete and public memory is uneven. Allowing the media layers to remain structurally independent reflects this fractured condition and resists the imposition of coherence where none exists.

Although the components are intentionally decoupled, their simultaneous presence enables moments of perceived alignment. The BPM-based timeline in TouchDesigner introduces a latent rhythmic grid, whereas the electroacoustic synthesis produces periodic or quasi-periodic behaviours as a result of grain structures and modulation patterns. These independent processes may occasionally appear to coincide, generating brief perceptual synchronies that emerge not from design but from the contingent interaction of autonomous systems (Figures 5 and 6).

This phenomenon parallels the behaviour observed in Ligeti's *Poème Symphonique*, in which individual metronomes begin independently yet occasionally fall into transient alignment. In the present work, such moments underscore the difference between structural independence and perceptual coherence.

3.3. Role of the performer

In the performance, I occupy a mediating position within the work. Seated at a table with a microphone and computer, I deliver the text live while operating the SuperCollider system. My role involves both narration and interaction with the electroacoustic process, yet it does not extend to controlling the procedural visuals or the projected text. This creates a performance situation in which I am embedded within, but not fully authoritative over, the surrounding media environment.

My manner of vocal delivery is flexible and responsive. Although the projected text unfolds autonomously, my spoken narration introduces variation in pacing, intonation and timing, which in turn shapes the behaviour of the concatenative synthesis. My presence therefore anchors the temporal experience of the work without imposing order on the independent visual systems. This mode of performance recalls, for me, the role of a dalang in Javanese wayang, who mediates between multiple layers of action while maintaining an interpretative rather than purely directive function.

At the same time, my spoken and projected text operate as a kind of instruction set for perception. Rather than issuing commands to the system, the narration offers cues that guide how audiences organise what they hear and see, indicating which

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

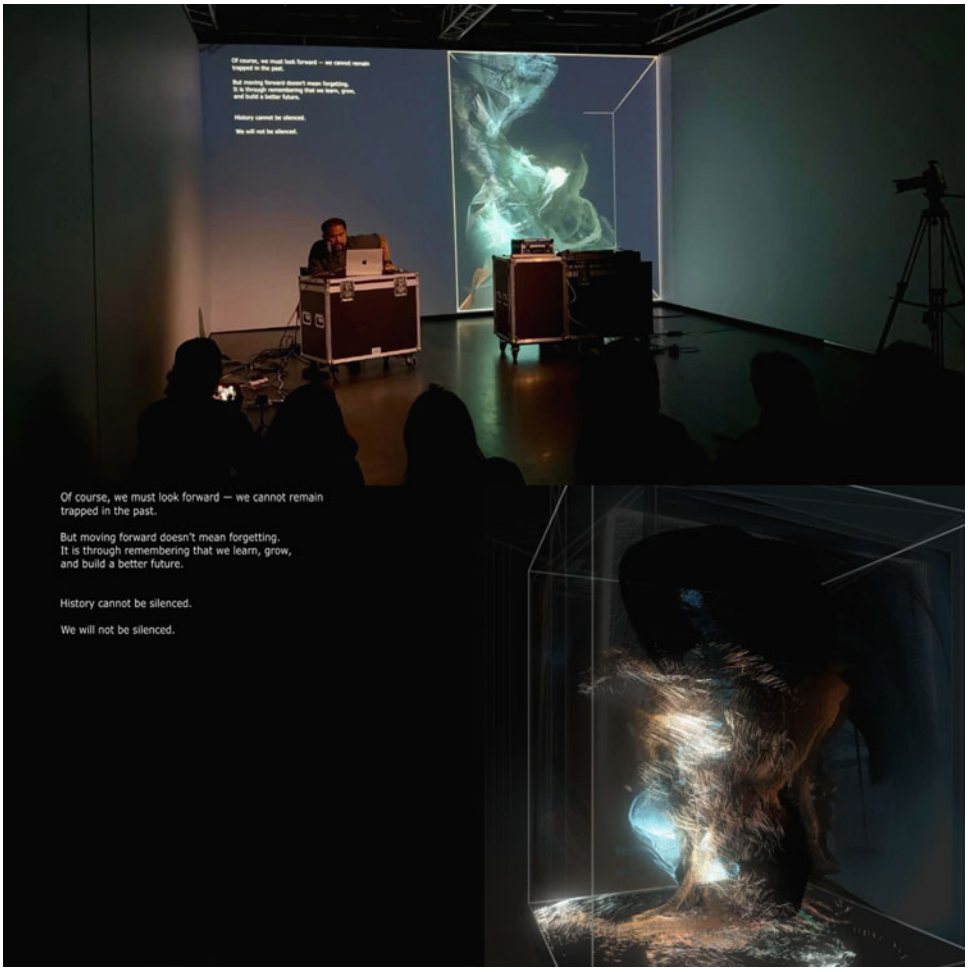


Figure 4. 'F2' Live Performance and image capture from performer screen.

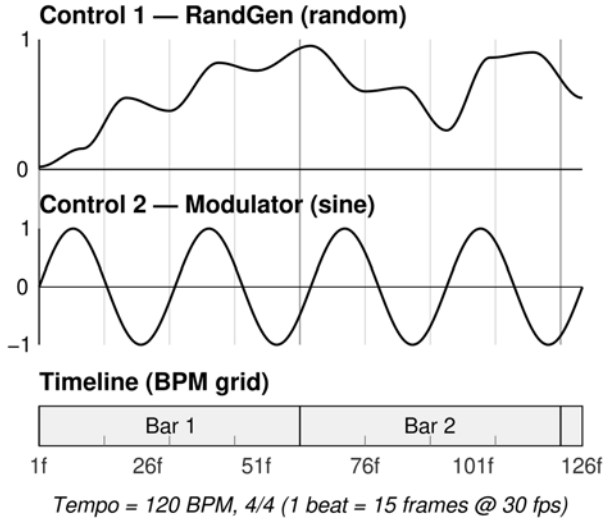


Figure 5. Two signal-based automation controls (above), whose behaviour is temporarily aligned with the timeline (BPM) below.

aspects of the electroacoustic texture or procedural visuals might be read as particularly charged or connected to specific historical references. In this respect, my role is less that of a controller of media and more that of a mediator of attention and interpretation (Figure 7).

3.4. Narrative material

In the projected and spoken text, I incorporate historical background on the May 1998 riots and debates on historical revisionism, together with the account of Fransisca, an 11-year-old victim documented in journalistic and human rights sources. As a Chinese Indonesian whose family, including my wife, was directly affected by the violence, I am close to this history; however, I do not pursue an autobiographical narrative here, as such testimony lies beyond the scope of this article. Instead, I treat the narrative as a source of thematic and affective orientation; the text provides context, frames perception and introduces emotional inflection, but it does not dictate or synchronise the behaviour of the audiovisual layers.

I present these materials in a documentary-informed manner to emphasise the tension between representation and abstraction. The procedural visuals present traces of the documentary footage without reproducing it; the electroacoustic sound transforms the spoken voice; and the text offers referential clarity while remaining structurally autonomous. In combination, these elements explore how memory, testimony and trauma circulate across media without collapsing into a single narrative.

Given that the material concerns sexual violence against a child and collective trauma, I am aware of the ethical risks. My strategy is to avoid graphic depiction and instead combine audiovisual abstraction with the centrality of text, so that the traumatic content is framed through language and context rather than visual shock.

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

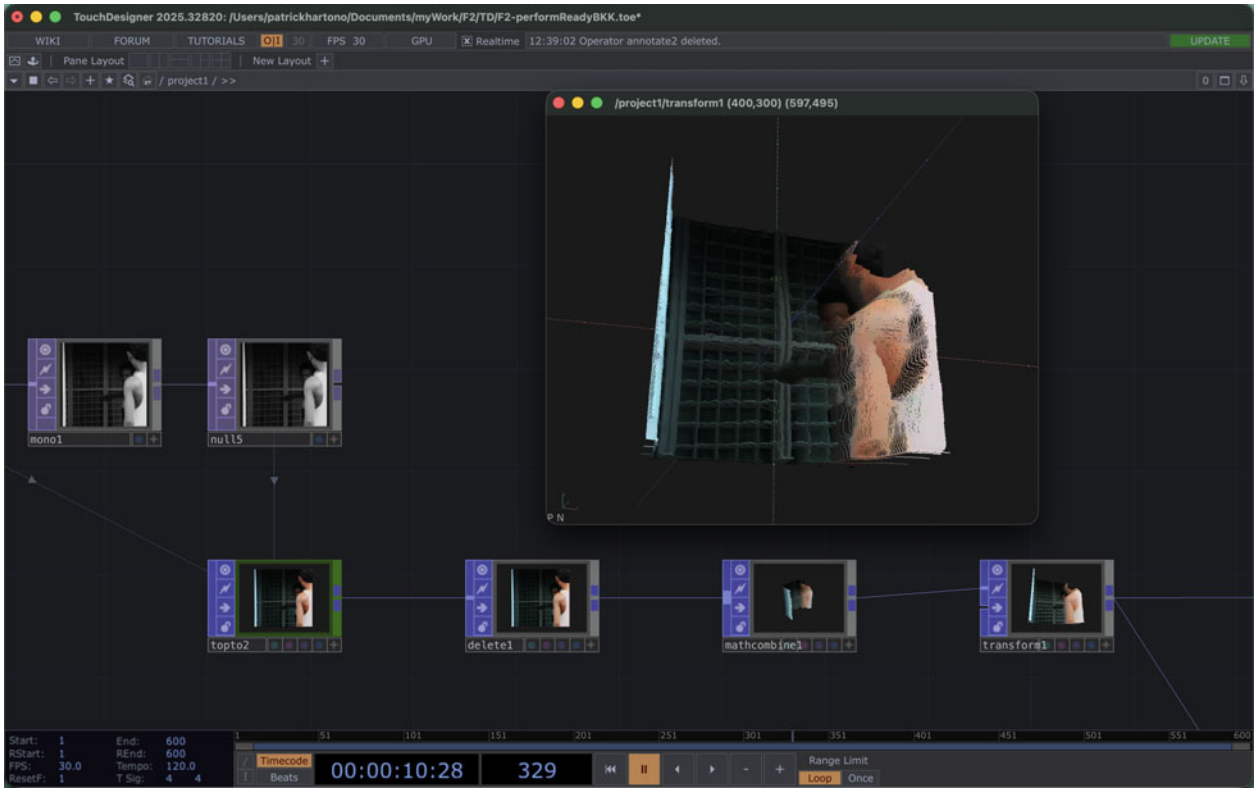


Figure 6. The procedural visual system in TouchDesigner: the pixel-to-geometry network, the resulting instanced geometry and the BPM timeline that drives it independently of the audio and text layers.



Figure 7. The performer mediates attention and interpretation from behind a laptop, shaping the audience's experience through care and distance.

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

The aim is to acknowledge the gravity of the events while maintaining a degree of distance. This ethical orientation underpins the work's unsynchronised structure and informs the way text, sound and image are positioned in relation to one another.

4. Audience study: method and findings

This study was motivated by audience reactions observed immediately after the performance. Several attendees approached the performer visibly emotional, and a number of them were crying. These unsolicited responses indicated that the work had generated a strong affective impact and led me to conduct a more structured audience study.

4.1. Methodology

I used an online questionnaire (Google Forms) consisting of seven closed items (multiple-choice and checkbox) and seven open-ended questions. The closed items asked about nationality/ethnic background, prior knowledge of the May 1998 riots, preference for anonymity or willingness for follow-up contact, which elements of the performance affected participants most, whether they read the projected text, whether the text helped or distracted from the experience and how they perceived the relationship between sound and visuals and between sound, image and text. The open questions invited short written reflections on the sound/music, perceived interactions between media layers, personal connection to the story, memorable moments, overall emotional response and any additional comments. The form opened with a brief explanation of the project, stated that participation was voluntary and confidential and asked for consent; respondents could choose to remain anonymous or leave an email address for possible follow-up only. In the analysis, all data were treated in anonymised form.

I circulated the questionnaire one week after the performance to allow time for reflective recall rather than immediate post-event reactions. Eleven participants completed it, roughly half of those who attended. Respondents included Chinese and other East Asian, European, Indonesian and American participants with varying familiarity with the May 1998 riots, from well informed to previously unaware, providing a diverse interpretative context. I analysed the material using manual thematic coding: reading all responses several times, noting recurring descriptions of text, sound, visuals and emotion and grouping them into broader themes concerning perceived media relationships, affective intensity and historical framing. Given the exploratory and practice-based nature of this study, the audience responses are not intended to be generalisable, but to illuminate how meaning-making operates within this specific electroacoustic audiovisual configuration. The full questionnaire is available from the author upon request.

4.2. Thematic findings

In reviewing the responses, I identified three recurring themes. First, **text as interpretive anchor**: participants repeatedly described the projected text as central to how they made sense of the work, often naming it as the element that allowed them to understand and emotionally process what they were hearing and seeing. Second, **emotional intensification**: across different backgrounds, respondents reported strong affective reactions, including intensity, discomfort, sadness and a sense of urgency, with some describing the experience as at times overwhelming. Third, **emergent coherence in an unsynchronised setting**: despite the absence of formal synchronisation, many participants perceived the three layers as coherent, describing sound, image and text as mutually reinforcing rather than fragmented. A smaller number of responses also expressed moments of confusion or difficulty in situating the traumatic material, indicating that the same configuration could be experienced as both powerful and challenging.

4.3. Key data excerpts

Representative comments included descriptions of the sound as 'overwhelming like a roar of rage', the text as 'the thread I needed to follow' and the visuals as 'metaphorical representations of the emotions connected to the sound and text'. One participant wrote that 'the text was the fulcrum of meaning of the whole audiovisual experience', and another noted that the elements 'carried different weights but interacted powerfully'. Alongside these accounts, some respondents indicated that the density of sound and the gravity of the narrative could be exhausting, suggesting that the work's emotional force and use of traumatic historical material were experienced as demanding as well as engaging.

5. Analysis: text achieving added value

5.1. Text as interpretive catalyst

From the audience responses, I found that the projected text functioned as the primary catalyst for interpretation. Participants described the sound and visuals as intense, but reported relying on the text to understand what the work addressed and to organise emotional responses. Some characterised the text as 'the thread [they] needed to follow' or 'the fulcrum of meaning of the whole audiovisual experience'. In this sense, the textual layer did not supply supplementary information but initiated the process through which the other media became legible as an engagement with historical violence and its aftermath, consistent with Ryan's account of narrative cues across media.

This catalytic role was evident among respondents who had limited or no prior knowledge of the May 1998 events. For these viewers, the narrative about state violence, gendered trauma and historical revisionism provided a framework that transformed the electroacoustic sound from abstract intensity into a representation of rage, grief or urgency and reframed the procedural visuals as metaphorical rather than purely formal. The text thus acted as a mediating element that allowed audiences to relate sound and image to a shared conceptual field.

5.2. Achieving added value in unsynchronised composition

Within Chion's terms, added value refers to the way sound enriches the image, creating an impression of heightened meaning. In this work, I read the audience accounts as suggesting a different configuration: text enriches both sound and image within an unsynchronised structure. First, the text amplifies the perception of sound. Participants repeatedly described the music as intense, overwhelming or 'like a roar of rage', and some explicitly linked this impression to knowledge of the narrative. Awareness of the child victim's story and the broader context of anti-Chinese violence appeared to increase the emotional weight of sonic gestures that might otherwise have been perceived as abstract or merely loud.

Second, the text reframes visual abstraction. The procedural visuals were widely interpreted as metaphorical representations of emotional states or historical turbulence rather than as autonomous

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

digital forms. Respondents described the images as 'illustrating and supporting the felt content of the piece', suggesting that textual knowledge of trauma and suppression encouraged them to read shifting geometries as traces of bodies, crowds or unstable environments.

Third, text provides emergent coherence across the unsynchronised components. Despite the absence of technical linkage, many participants reported that sound, visuals and text worked together and 'interacted powerfully'. Coherence, in this case, appears not as a result of structural alignment but as a perceptual effect produced when the narrative layer allows audiences to infer relationships among independently evolving media streams. Text thereby becomes a generator of added value that operates across, rather than within, conventional synchronisation regimes.

5.3. Narrative cueing and crossmodal association

In my analysis, these observations can be related to two perceptual mechanisms. Narrative cueing is evident in the way textual milestones – shifts from contextual information to the detailed account of the victim's experience, or from description to reflection – coincide with reported changes in how participants listened. When the narrative darkens, respondents describe the sound as more oppressive or urgent, even though there is no programmed linkage. Crossmodal association follows: guided by these cues, viewers relate specific sonic and visual events to fragments of the narrative, constructing correspondences that are not encoded in the technical system. Added value, in this configuration, is therefore produced through perceptual integration and interpretation rather than through temporal coincidence or signal-level mapping.

5.4. Triangular meaning-making model

Taken together, I understand the findings as supporting a triangular added-value model (Figure 8) in which sound, procedural visuals and text occupy distinct positions and interact through asymmetric perceptual relations. Unlike a generic three-element diagram, this model is directional; text functions as the primary mediator, whereas sound and image are positioned as recipient-modulators whose perceptual reading is shaped by textual cues. The model is not a claim about how unsynchronised audiovisual works function in general; it is a description of the specific perceptual dynamics observed within F2, articulated in a form that may inform the analysis of other configurations. Three pairwise relations populate this triangle, and I describe them in turn.

The first relation runs from text to sound. As described in Section 5.2, audience accounts indicate that knowledge of the textual narrative shapes how the electroacoustic sound is perceptually categorised. Sonic gestures that might otherwise have been read as abstract intensity or noise were repeatedly described in affective terms – as 'rage', 'grief' or 'urgency' – once participants had access to the textual content. The text therefore performs a semantic anchoring function: it does not modify the sound itself, but it directs perceptual attribution by supplying a referential frame within which sonic events become legible as expressive of particular affective states.

The second relation runs from text to image. Participants consistently described the procedural visuals as metaphorical representations of bodies, crowds or unstable environments rather than as autonomous digital forms. This metaphorical reading was clearly contingent on the textual narrative; the same shifting geometries, presented without text, would not have been read in this way. The text here performs a reframing function; it does not alter the visual material, but it routes perception towards figurative readings that the procedural abstraction alone neither denies nor specifies.

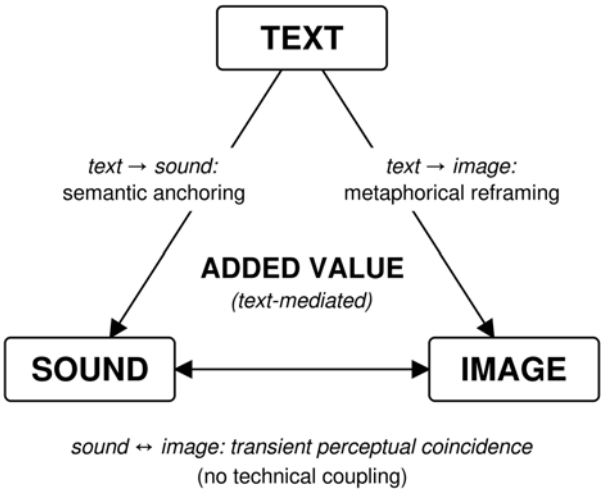


Figure 8. A triangular added-value model based on F2. Text functions as the primary directional mediator; it semantically anchors sonic intensity (text → sound) and metaphorically reframes visual abstraction (text → image), whereas sound and image relate only through transient perceptual coincidence (sound ↔ image), without technical coupling. Added value emerges from this asymmetric configuration through textual mediation, rather than from direct sound–image interaction.

The third relation, between sound and image, is markedly different. Because the two layers share no technical coupling, no parameter mapping and no scripted alignment, their relationship is not directional but contingent. Audience reports of perceived coherence between sound and image arose not from causal linkage but from transient perceptual coincidences – moments at which independent rhythmic or textural events were retrospectively heard as connected. This is the Ligetian parallel discussed in Section 3.2; independent processes can fall briefly into apparent alignment without being structurally coupled. Within the model, sound–image coherence is therefore an emergent rather than constitutive relation, and its perceptual stability depends on the textual frame that the other two relations supply.

Read together, these three relations suggest that what Chion calls added value, in the configuration examined here, is generated not by sound enriching image (the original cinematic formulation) but by text directionally enriching both. Sound and image become jointly significant – as material that addresses historical trauma and its contemporary erasure – through textual mediation rather than through their own coupling. The triangular configuration thus extends Chion's concept into a context of deliberate compositional unsynchronisation, while remaining specific to the conditions of F2; the model describes how text can function as a primary directional mediator within one such configuration, not how unsynchronised audiovisual works function in general.

6. Discussion

6.1. Implications for audiovisual composition

The findings suggest that text can operate as a fully legitimate compositional layer within electroacoustic audiovisual practice, rather than as an ancillary or explanatory device. In the work discussed here, textual narration does not simply accompany sound and image; it shapes how both are perceived, providing a

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press

conceptual and affective framework through which audiences understand the performance. This indicates that, in multimodal composition, the distribution, pacing and structuring of text can be treated with the same level of attention as timbral design, spatialisation or visual form. Text can thus be integrated into compositional thinking as time-based material that participates in the organisation of the work's overall trajectory.

6.2. Implications for new media performance

Within generative and real-time electronic performance, the case study highlights how text can coexist with autonomous systems without being reduced to a control signal or interface label. In this configuration, projected text and live narration function as points of human-readable orientation within a field of algorithmic behaviour, mediating the audience's engagement with processes that remain partially opaque. At the same time, the use of explicitly traumatic historical material underscores the need for ethical attention to content warnings, framing and the affective demands placed on audiences. New media performance that incorporates such material must balance the capacity for immersion and intensity with an awareness of care, distance and the role of the performer as mediator. Read across these implications, the study also extends Chion's notion of added value beyond its cinematic origin, situating it within contemporary intermedial performance, where live presence, generative systems and unsynchronised structures are central.

7. Conclusion

This article has examined an electroacoustic audiovisual performance in which sound, procedural visuals and text are deliberately left unsynchronised, yet were experienced by audiences as emotionally coherent. The audience study indicated that the projected and spoken text served as a central interpretive anchor, amplifying the perceived intensity of the sound and reframing visual abstraction as metaphorically connected to historical violence.

The analysis proposed that this configuration can be understood through a triangular added-value model with three pairwise relations: text semantically anchors the perception of sound, text metaphorically reframes the perception of image and sound and image themselves relate only through transient perceptual coincidence. Text thus functions as the primary directional mediator. This contributes to both compositional practice – by affirming text as a viable structural layer in electroacoustic audiovisual work – and to theory, by extending Chion's concept of added value into the domain of live, unsynchronised performance, while remaining grounded in the specific case of F2.

Future research might explore how larger or more diverse audience studies could refine understanding of cross-cultural responses to multimodal representations of historical trauma, and how automated or generative text systems might further develop this triangular model.

References

Bachfischer, G., Robertson, T. and Zmijewska, A. 2006. Typography in Motion: A Framework of Moving Type Use'. *WSEAS Transactions on Information Science and Applications* **3**(10): 1810–17.

Basanta, A. 2015. Extending Musical Form Outwards in Space and Time: Compositional Strategies in Sound Art and Audiovisual Installations. *Organised Sound* **20**(2): 171–81. https://doi.org/10.1017/S1355771815000059.

Blackwell, A., Cocker, E., Cox, G., McLean, A. and Magnusson, T. 2022. *Live Coding: A User's Manual. Software Studies*. Cambridge, MA: The MIT Press.

Brownie, B. 2015. *Transforming Type*. London: Bloomsbury. https://www.bloomsbury.com/ca/transforming-type-9780857856333/.

Carvalho, R. 2025. Sound-to-Visual Mapping Strategies in Audiovisual Performance and Installations. In L. C. Castilho, R. S. Dias, L. Rocha and A. De Sousa Dias (eds.) *Perspectives on Music, Sound and Musicology II*. Cham: Springer. https://doi.org/10.1007/978-3-031-67503-4_21.

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. Doctoral, Bournemouth University. https://eprints.bournemouth.ac.uk/40772/.

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

Emmerson, S. 2007. *Living Electronic Music*. Aldershot: Ashgate.

Hill, A. 2013. *Interpreting Electroacoustic Audio-Visual Music*. http://hdl.handle.net/2086/9898.

Ikeshiro, R. 2012. Audiovisual Harmony: The Realtime Audiovisualisation of a Single Data Source in Construction in Zhuangzi. *Organised Sound* **17**(2): 148–55. https://doi.org/10.1017/S1355771812000076.

Luthfiasari, A. and Fadillah, C. 2023. 'Comparative Analysis of Kinetic Typography Elements in Movie Title Sequence'. *Serat Rupa Journal of Design* **7**(1): 43–60. https://doi.org/10.28932/srjd.v7i1.5636.

McPherson, A. and Tahıroğlu, K. 2020. Idiomatic Patterns and Aesthetic Influence in Computer Music Languages. *Organised Sound* **25**(1): 53–63. https://doi.org/10.1017/S1355771819000463.

Ryan, M. 2004. '*Narrative across Media: The Languages of Storytelling*'. Lincoln, NE: University of Nebraska Press.

Spence, C. 2011. 'Crossmodal Correspondences: A Tutorial Review'. *Attention, Perception, & Psychophysics* **73**(4): 971–95. https://doi.org/10.3758/s13414-010-0073-7.

Spence, C. and Parise, C. V. 2012. The Cognitive Neuroscience of Crossmodal Correspondences. *I-Perception* **3**(7): 410–12. https://doi.org/10.1068/i0540ic.

Tsang, W. S. 2023. The Cartographies of Place: Approaches to Audio-Visual Composition Incorporating Aspects of Place'. Edith Cowan University. Application/pdf. https://doi.org/10.25958/Y35T-2940.

Villaseñor-Ramírez, H. 2023. Live Coding Outside, Live Coding Inside: Listening, Participation and Walking. *Organised Sound* **28**(2): 195–205. https://doi.org/10.1017/S1355771823000353.

Wishart, T. 1996. *On Sonic Art*, ed. S. Emmerson. Rev. edn. Amsterdam: Harwood Academic.

Xambó, A. and Roma, G. 2024. Human–Machine Agencies in Live Coding for Music Performance. *Journal of New Music Research* **53**(1–2): 33–46. https://doi.org/10.1080/09298215.2024.2442355.

Zainuddin, M. Z. M., Din, S. C. and Lazim, N. A. M. 2025. Exploring the Role of Typography in Documentary Videos: A Case Study on Mineral Resources'. *Idealogy Journal* **10**(2): 164–71.

Zareei, M. H. 2020. Audiovisual Materialism. *Organised Sound* **25**(3): 362–71. https://doi.org/10.1017/S1355771820000321.

https://doi.org/10.1017/S1355771826101605 Published online by Cambridge University Press