INTERNATIONAL JOURNAL OF FILM AND MEDIA ARTS (2026) Vol. 11, Nº. 1 pp. 94-131

ijfma.ulusofona.pt
doi: 10.24140/ijfma.v11i1.10693

FROM PANEL TO PULSE: AURALITY, CLOSURE, AND READER AGENCY

CHRIS LAMBERT
SAE University College, Australia

A musician, producer and educator in creative media for 25 years, **Christopher** has taught at Flinders University, SAE University College, and TAFE campuses. Holding a Master of Creative Industries, he currently lectures in the Bachelor of Audio program at SAE University College.

Corresponding author
Chris Lambert
chris.lambert@sae.edu.au

Schedule for publication
Paper Submitted: 7th October 2025
Accepted for Publication: 28th January 2026
Published online: 30 May 2026

Abstract:

This paper examines the integration of audible sound into the traditionally silent medium of comics, analysing how emerging technologies reshape narrative structure, reader agency, and sensory experience. Drawing on theories of closure, intermediality, and visual semiotics, it evaluates historical and contemporary forms including motion comics, Webtoons, AR/VR comics, and audio comics. The study argues that synchronised or prescriptive sound often disrupts reader-controlled pacing and imaginative participation, undermining the core mechanics of comics reading. In contrast, fluid, atmospheric, and non-diegetic sound design can enhance immersion while preserving interpretive openness. Case studies demonstrate that successful sonic integration occurs when sound complements rather than dictates visual rhythm. The paper concludes that sound should function as an augmentative layer within comics, the notable exception being the fully sonic 'audio comic' which operates as a distinct mode of narrative construction.

Keywords: audible comics, aurality, digital comics, intermediality, closure, Webtoons

Reading traditional print comic books has never been a silent activity, nor is it a silent medium. The characters speak inside our minds in imagined timbres and tones. Noises travel off the page in onomatopoeic letter forms or traverse and fluctuate in dynamic lines and shapes. Radio speakers blast music through jagged lyric bubbles and floating notation. Clouds and stars cause us to experience the percussion of explosions and their ringing aftershock. In other words, comics utilise visual signs and symbols that semiotically refer to sound. This paper seeks to determine what happens when audible sound enters an acoustically silent tradition and how emerging technologies have informed and expanded this incursion.

In this paper, I conclude that unless a comic is remediated entirely sonically, audible sound is problematic when used as the primary method of directing narrative or dictating the pace of reading. Aurality in comics should complement both the semiotics of imagined sound and the visual rhythm of the work, allow a degree of traditional 'closure' and provide some freedom in the way in which the reader constructs narrative. This maintains the appeal of newer 'comic' forms to the traditional readership and positions them as part of the comics tradition, whilst making use of the transformative power of sound afforded by these new evolutions of the comic book.

I do not contest assertions that the visual comic "is an art form that is already complete" and I concur that "any reader looking at a comics page will imaginatively compensate, easily and spontaneously, for the absence of real motion or sound" (Groensteen, 2011, pp. 70–71). Comics with sound can be seen as a form of remediation, where one medium is seen as culturally reforming or reinventing another (Bolter & Grusin, 1999). Hague (2011) provided an example of sonic remediation of the comic form back as far as 1945, when New York City Mayor Fiorello La Guardia read and described newspaper comics on the radio during a newspaper deliverymen's strike. One could also consider the radio series dramatizations of *Buck Rodgers* and *Flash Gordon* from the 1930s in a similar light, given that their comic strips pre-dated them.

Although, their very nature determines that their legitimacy as true 'comics' will be contested, I document and assess the use of sound in evolutions of the comic form, including digital and motion comics, Webtoons, VR/AR comics, external soundtrack for printed comics and the phenomena of 'audio-comics'. As addendum, I also put forward my own speculative proposals for use of the immersive properties of voice cloning in narrative, and for a method of remediating

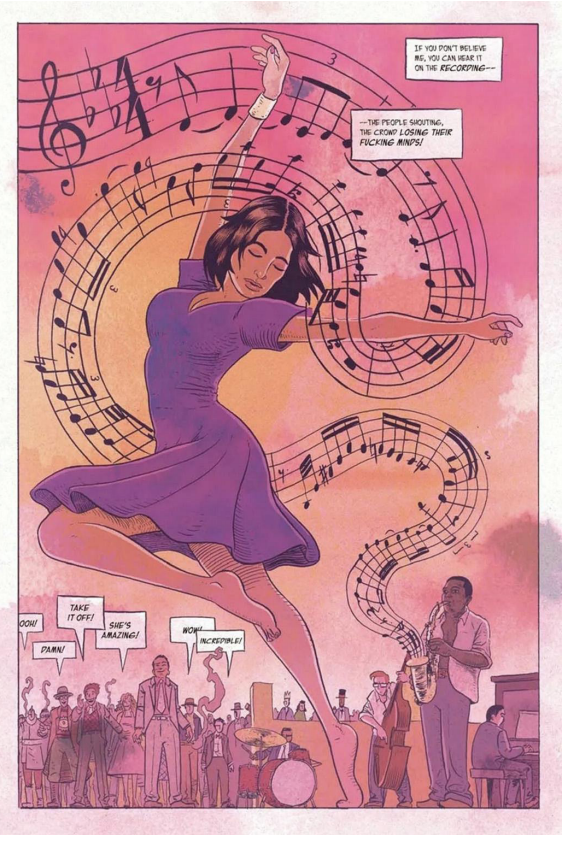

FIGURE 1 One of the many creative ways to visually represent sound—from *Chasing the Bird* by David Chisholm (© Z2 comics, 2020, reproduced with permission of the artist).

print comics to sound, employing the auto-creation of immersive soundtrack to accompany narration for the BLV (Blind and Low-Vision) community using MLLMs (Multimodal Large Language Models).

Closure, imagined sound and aurality

Comics and graphic novels provide perhaps one of the most creative and varied lexicons for actualising imagined sound. Pellitteri's extensive interrogation of the multitude of visual sound conventions and devices brought to bear in the medium speaks to this (2019). Even so, as technology has brought cheaper, more ubiquitous, and more efficient audible sound reproduction into everyday use, it is natural for us to explore ways in which aurality can inform the comic book experience, not just as an accessory, but as an intermedial transformation, capable of redefining reading, listening and narrative immersion.

Empirical studies specifically about reader reception of new comic forms with audible sound and of soundtracks used with traditional print comics are essentially non-existent. The best available evidence is in case studies of digital/motion comics and empirical work on Webtoons, informed by works that attempt to define the comic book and the reader experience. Theorists such as Groensteen have ventured that sound and vision, whilst vital to cinema, will in the context of comics, "never be more than add-ons" (2011, p. 71). However, the work of Ian Hague created a new platform of acceptance for comics with audible sound by critically engaging with its potential and application in *Comics and the senses* (2014).

Postema (2013) makes a solid argument that narration in comics is an image-based process, a visually driven form to which the verbal often adds interest or depth but is not necessary. With the long tradition of comics as primarily visual medium including the visual representation of sound, it is worth examining how comics are processed and understood via the concept of closure and how physical sound disrupts this paradigm.

The concept of closure involves the mental leap that readers make between panels to perceive a continuous narrative. Whilst the term, when used in the context of comics, belongs to Scott McCloud (1993), the foundation of this concept of the reader's role in narrative was laid by Gestalt psychologists who posed the principle that perception tends toward the completion of incomplete forms (Koffka, 1935; Wertheimer, 1923); by art historian David Kunzle (1973) who wrote about how sequential narrative depends on reader inference; by semiotic analysts (Eco, 1979; Fresnault-Deruelle, 1976) who explored reader cooperation and inferential gaps; by artist Will Eisner (1985) and scholar Benoît Peeters (1991) who explored this in relation to the comic panel; and the concept was extended upon by theorists such as Thierry Groensteen (1999/2007) and Neil Cohn (2010). In comics there is a visual fragmentation of time and events, and much is filled in with the reader's multisensory imagination, including the sonic dimension that exists only through abstraction.

In Benoît Peeters' (1991) theory of *mise en page,* the panel is compared to both painting and film and found to be neither, rather, sitting somewhere in-between the two. For Peeters the reader does not simply 'close' the 'gaps' between panels, but instead interprets the impact of the page as a whole, including the influence of reader memory and the *decoupage*—a term borrowed from French film theory encompassing the panel layout, and including the organisation, relative scale, direction and all the other visual cues that can carry [or disrupt] a narrative, inside a 'phantom panel' that exists in the reader's mind.

Peeters provides an important and oft-referenced framework for examination of how comics are constructed, and it is significant here as visually represented sound becomes imagined sound and can persist via his 'phantom panel'. His work refers to the mental construction between the frames by the reader

as "off-screen", a "mental image that persists after reading" (Mougin, 2023). Peeters quoted Art Speigelman's insight "cartooning is the art of turning time back into space" (The Lakes International Comic Art Festival, 2020), and when time is mediated into the visual by comic creators the interpretation of temporal aspects, such as sound, will vary with the reader.

No matter how we classify a layout, the way a comic will be read and interpreted is wholly dependent on the individual experience, characteristics, preferences and even mood of the reader. Cohn (2021) dedicates an entire work to examining visual language processing, affirming that there is no high level of uniformity in the way that individuals read and comprehend the same information. When discussing closure, McCloud (1993) mentions media analyst Tony Schwartz, who speaks of how "the audience responds to the message in the context of their own lives" (1983, p. 3). Cohn advances on this, in examining the processing of visual communication in relation to an individual's cognitive *idiolect*, their language framework of understanding that "developed from the patterns they are exposed to and practice with across their lifespan".

Further, Cohn recognises that this sequential image comprehension that is the "union of conceptual information that is grouped via unconscious hierarchic structures in the mind" (2010, p. 128), is not unique to visual narrative, and "resemble those involved with sequencing in language and music" (2021, p. 55). Building on the social aspects of this collaboration with the reader's mind, Adler (2011) eloquently relates that in the silence of comics, not only is the reader "given the opportunity to fill in the details that are missing between the frames" but is given the opportunity to respond "with an intense emotional, intellectual and/or critical reaction to what is not articulated explicitly and [is] therefore restored through his/her own understandings" (p. 2279).

Whilst Peeters (1991) recognises the importance of reader agency in interpretation, he does attempt to classify four styles of comic layout based on the struggle between image and narrative. In his typology, layouts are labelled *conventional, decorative*, *rhetorical* or *productive*. These categories are explored in depth by Groensteen (1999/2007), who finds them insufficient, in that, many layouts can fall within two or more of them simultaneously.

Groensteen's (2012) more holistic 'spatio-topical' system transforms McCloud's 'closure' "into a core element of comics functioning" (Trabado, 2022, p. 243). It considers the engagement of all senses, the locale not just on the page, but in the entire work and separate analysis of space and place.

Whilst Groensteen (1999/2007) speaks of closure in his *System of Comics* in terms of the frame—as the enclosure of a "fragment of space-time"— he is also addressing the audience's ability to close imaginative gaps, speaking of the fusion of "documentary drawings" and "imaginary drawings" (pp. 55–71). Groensteen's spatio-topical system gels with Cohn's and Adler's emphasis on the role conceptual memory plays in closure and affirms that closure could take place between a panel held in memory and any subsequent panel, on another page, or even across volumes of work. This persistence of memory in such leaps, operates easily with imagined sound, but is at odds with the linear nature of audible sound.

In these semiotic frameworks all visual sound elements rely on reader interpretation. Groensteen in his early work identified comics as a "predominantly visual medium" (1999/2007, p. 189) and examined the importance of visual sound, captions and other verbal information in its relationship to the image. If one is to shift these features into the aural domain, the visual abstraction of the auditory imagination is lost and moves comic reading into an intermedial assemblage. Groensteen (2011) sees this as disruptive to the comic experience, concluding that "rather than strengthening the hold of the fictional world, [the effect of multimedia] is to destabilize that world or to transform it into something else" (p. 71).

Comics as intermedial assemblage

Unlike McCloud (1993) who locates the symbolic purity of comics in a system of signs that is solely visual, Hague (2014) recognises greater possibilities for the intermedial aspects of comics in his work. He puts aside the notion that physical sound cannot be a part of comics, citing social definitions from Barker (1989), who believes that "a comic is what has been produced under the definition of a comic" (p. 8), and Witek who asserted "to be a comic means to be *read* as a comic" (cited in Heer & Worcester, 2009, p. 149).

In these more expansive definitions of comics, Hague (2014) identifies four main categories of how aurality can be experienced: sounds **of** comics (for example page rustles and turns); sounds **in** comics (part of the comics narrative and produced by the comic object, for example digital comics); and sounds **with** comics (packaged with the comic, but reproduced by external means, such as a vinyl record or suggested playlist as outlined later in this paper); and even sound **as** comics.

Hague (2014) successfully exhausts the possibilities of the sounds **of** comics, and so this paper focuses on experiments with sounds **in** comics (Marvel CyberComics, DC Motion Comics, MadeFire Motion Books, WebToons, VR Comics); sounds **with** comics (BLV readers, AR and VR comics, and Music soundtrack); and sound **as** comics, where a comic work is totally remediated aurally.

The sonic dimension's effect on closure

The linear nature of added synchronised soundtrack resists the agency of the reader in temporal flexibility. Their ability to pause, linger, re-visit, jump, reverse or accelerate in the narrative is greatly affected. Baroni's chapter in the book *Unnatural Narratology* (2020) speaks to the non-linear nature of comic book narrative, and the reader mediated consumption of events on a page.

The silence of comics is what activates the imagination most vividly, with multisensory closure in which, our imaginations invent the soundtrack as we read. Idhe (2008) speaks of the freedom of interpretation in imagined sound as being "changeable and variable" (p. 122), but when we provide aurality this aspect of multisensory closure is replaced, as is our freedom to interpret, reader agency replaced, as it were, by a new framework of authorial control.

However, whilst introducing the sonic dimension does push the comic form towards the cinematic, it does not destroy closure entirely and it provides new narrative opportunities in a multimodal negotiation. As the reader reconciles visual gaps, soundscapes can allow new dimensions of authorship, whilst still leaving some closure opportunities for the interpretation of the provided sound. Hague (2014) affirms this by positively positioning audible sound as a tool with which creators can negotiate between producer-intent and receiver-input and renegotiate co-construction in this contract.

Looked at from a Deleuzian perspective this coupling becomes a new assemblage—the combination of the comic's visual rhythm and the imposed sonic rhythm. Audible sound becomes a 'foreign body' in a mono-sensory visual system, operating autonomously, negotiating the intermedial interstice between the frozen temporality of the visual and the unfolding duration of sound, resulting in a framing of sound (*cadrage sonore*), where the reader must mediate between the crosscurrents of the comic's reader-controlled time, and sound's linear author-controlled time (Deleuze, 1989).

First Steps: Marvel's CyberComics

An early example of the requirements of this kind of mediation is with Marvel *CyberComics*, a short-lived initiative on Marvel's website in 1996. These interactive comics, built with Macromedia Shockwave authoring software, had soundtracks, voice acting, and sound effects embedded into

panel transitions. These works marked one of the first large-scale attempts by a mainstream publisher to 'graft' sound onto comics' traditionally silent visual grammar, with sound and image operating as parallel but non-identical registers.

Readers of *CyberComics* were compelled to negotiate between the visual-temporal logic of comics (frozen images, spatialized sequence, reader-controlled pacing) and the sonic-temporal logic of digital media (continuous time, author-controlled pacing). The concept of closure, therefore, does not disappear, but migrates to a new level: the reader now performs closure not only between panels, but also between modalities, reconciling the imaginative gaps of the visual with the real-time flow of sound.

Sometimes the two logic systems align. Sometimes they align seamlessly; at other times, even with a loosely synchronised sound, the artificial quality of a poorly chosen or designed sound jars against the imagery, producing cognitive dissonance. The clash between auditory and visual sound cues is particularly disruptive. One example of this dichotomy is where automatic doors close behind Professor X in the X-men *CyberComic*; the onomatopoeic visual sound flashes up behind him in a dramatic "TWOM", while the ill-timed sound effect has all the impact and gravitas of the closing of a wooden kitchen cupboard.

The reader must actively negotiate this gap, performing a kind of second-order closure in the interstice between sound and image. As with this instance, the gaps for closure in *CyberComics* are often large, as during this era bandwidth was low, Internet sound quality was poor, and sound appeared to be a secondary consideration.

Marvel's *CyberComics* project exemplifies both the promise and the problem of the inclusion of sound. On the one hand, it aligned with late-1990s enthusiasm for multimedia convergence, when comics, film, and games were increasingly imagined as mutually reinforcing transmedia forms in 'world-making' (Jenkins, 2006). Soundtracks and effects offered a way to render comics more cinematic, appealing to new audiences acclimated to audio-visual media. On the other hand, their overall effect undermined the autonomy of reading and the imaginative space afforded by silence (Hague, 2014).

If we consider closure in print comics in terms of reader autonomy, *CyberComics,* already limited by linearity due to their forward-back button navigation system, imposed a sonic rhythm in their looped soundtracks and sound effects that synchronized reading time to audio playback, thereby constraining agency in variable pacing. Hague (2014) does find one small boon in this, suggesting that, to a degree at least, aurality could also be employed as a tool to assert a certain degree of authorial temporal control.

This sonic rhythm may also conflict with the rhythm of the layout if we consider Groensteen's concept that spatial arrangements of comics can convey temporal significance. The complexity of this concept is unfolded in Mikkonen (2017) who, in applying the ideas of Russian formalist Gérard Genette to comics, mapped the duality of visual temporality via story-time (chronological events in the story world) and discourse-time (how those events are presented), mediated through three narratological dimensions: order—the arrangement of events relative to chronology (flashbacks, flashforwards); duration—the relationship between the time represented and the space used to represent it; and frequency—how often events are narrated.

It is through these dimensions that comics manipulate rhythm, pacing, and narrative emphasis. This temporal meaning emerges through multiple structural layers: the visual, including panel relations (for example the transitions and gaps between panels) as well as page layout and composition; the narrative sequencing; and reader inference and interpretation. Mikkonen's chapter hints at a key tension:

based as it is on literary works, Genette's model assumes narrative discourse unfolds linearly, whereas comics operate spatially. Adding sound to comics intensifies this tension in multiple ways.

With reader agency, narrative time is elastic, whereas the audible produces objective duration. Music has length, voice has duration and ambience unfolds over seconds, whereas a reader may scan a panel or pause on it and contemplate at will. Sound collapses spatial simultaneity as multiple panels can coexist spatially, but sound cannot be spatially frozen in the same way. This creates a design challenge for comics with sound, in that duration, rhythm, and sequence must find points of attachment. If sound accompanies a page, the creator must decide if each panel will have its own sound, if sounds will persist across panels, and whether sound will follow a set reading order.

Examining the act of adding aurality under Genette's framework brings focus to how this creates a hybrid of comics, animation, film and audio drama, and how the dual timeline of spatial narrative and fixed duration audio will often clash, causing friction between story-time, discourse-time and listening time.

FIGURE 2 Screengrab of the AOL Marvel Online promo CD-ROM. Marvel Characters Inc., 1997.

When sound displaces silence, the reader's role in producing auditory meaning through closure is mostly lost. Deleuze (1989), however, helps us see what is gained: an expanded aesthetic of intermediality, where sound and image resonate through their differences rather than their sameness. In this sense, Marvel's *CyberComics* can be understood as an early experiment in deterritorializing the comic form. By incorporating sound as a 'foreign body', they produced interstitial spaces where closure was not entirely eliminated but transformed. The challenge for contemporary practitioners is to embrace this interstitiality, not as a gimmick or cinematic mimicry, but as a unique aesthetic mode in which the silent rhythm of comics and the audible rhythm of sound move into dynamic, Deleuzian relation.

Motion comics

Following Marvel's experiments, Warner Brothers and DC introduced the term 'motion comic' in 2008, hoping to spark "a new kind of Web entertainment: a hybrid of comic books and animation" (McBride, 2008). Their first foray into this medium was the Batman series *Mad Love,* created as promotional content for the film *The Dark Knight* and was delivered for platforms such as Nokia phones, Xbox Live gaming consoles and iTunes. This series was quickly followed up by motion comics to promote the film *The Watchmen* (IGN, 2008).

Lars Ellström (2010) distinguishes between media-specificity (what a medium "does best") and intermediality (hybrid forms that blur boundaries). When sound and limited animation are introduced to comic's spatial-sequential medium they must sit in balance where the dominant medium is still leading the

FIGURE 3 *DC's Watchmen: The Complete Motion Comic*. © Warner Brothers Entertainment, 2009.

narrative. If executed poorly, we disrupt this balance, creating what Ellström calls qualified media products that operate with "fluctuating conventions", and retain one medium's logic while borrowing features of another (p. 57).

Bruhn (2016) interprets German media theorist Yvonne Spielmann as positing that the "mixing and transformation of conventional, distinct media forms" results in "intermedial products [that] stupefy and alienate media consumers and media users" (p. 2). Whether or not this was the intended interpretation, the way in which motion comics move media content into a new intermediality can result in the media and their associated significations being "radically dislocated and displaced" (Overell, 2014). Comic critic Chris Sims mused that motion comics resulted in the use of comic content "for an entirely different purpose than was intended, with predictable results" (as cited in Tucker et al., 2018, p. 21).

In the case of DC's motion comics, fan reception was ambivalent at best (Tucker et al., 2018), and at worst clashed with audience expectations (Morton, 2017), although much of the criticism of the adaptation was focused on the voice acting, concepts of framing panels and an overtly filmic portrayal (Brooker, 2023). However, Brooker (2023) does mention the clash of narration with the rhythm of the reader in the comic book and the unfortunate focus on background details due to spot sound effects.

Examined through the lens of Jacques Rancière's (2009) concept of the 'distribution of the sensible', comics distribute meaning through visual spacing and silent gaps, whilst motion comics redistribute the sensible order, by imposing sound as a guiding interpretant. In many fan discussions of motion comics, the guidance of sound was perceived as a loss of imaginative freedom, echoing the sentiments of Groensteen (2011) and McCloud (1993). As Morton (2015) posited, "our involvement as readers is rendered unnecessary by the inclusion of a soundtrack that performs the work for us" (p. 359). Speaking of the way in which sound is experienced from the printed page, the richness of this participatory experience is delineated by Peterson (2009) who muses that by "inviting the reader to experience the sound of an action or place is to encourage embodiment of space and give an animating force to the actions" (p. 164).

The digital comic book evolution: Madefire and Motion books

The first large-scale digital comic platform was ComiXology. It launched in July 2007 as an online community for comic book fans, where they could receive news of new comic book releases and develop pull lists (individual pre-orders) from 'brick-and-mortar' stores. Two years later, with the invention of digital tablet devices, a new platform called Comics by ComiXology was launched, a digital comic book reader and store for mobile devices, including iOS, Android, Windows 8, and an Internet web reader, all of which allowed users to access their digital comic collection across multiple devices.

It never offered digital-specific features such as sound, interactivity or motion and so could only distribute comics that were "reasonably compatible with print" and could be remediated as print comics (Kashtan, 2018, p. 127).

In sharp contrast, Madefire, launched in 2012 by Ben Wolstenholme, Liam Sharp, and Eugene Walden, was both platform and publisher, with aspirations to move beyond scanned-page digital comics. Its signature format, the Motion Book, introduced animated panel transitions, layered parallax, and integrated sound and multimedia, but generally used word balloons instead of voiceover (Allen, 2014). In the visual sense, with the viewer able to walk around scenes and transcend panel boundaries, Madefire was aligned with the trajectory McCloud (2000) envisioned in his 'infinite canvas' argument, wherein digital technology dissolves the fixed page into a spatially and temporally fluid field.

MadeFire's Motion Book tool, released publicly in 2012, enabled creators to embed multiple types of sound: ambient looping cues, such as weather or city environments, designed to frame the visual space; music cues for dramatic or emotional emphasis, closer to film or gaming conventions; and synchronous sound effects (for example, gunshots and footsteps) tied to panel transitions or interactive triggers.

As Aaron Kashtan (2018) noted, this signalled a shift from digital comics as "remediated print" toward a "native digital form" that embraced multimodal layering (p. 92). Sound thus operated not only as ornamentation, but as a structuring device, recalibrating pacing, and atmosphere.

Madefire raises important questions about the ontology of comics. Charles Hatfield (2005) has argued that comics rely fundamentally on the tension between the discrete and the continuous. By overlaying continuous sonic environments onto the discrete units of panels, Madefire blurred this tension in some comics and shifted the rhythm of reading toward something closer to audio-visual media. Similarly, Karin Kukkonen (2013) echoing sentiments of Groensteen and McCloud, emphasizes that comics should depend on readerly inference to fill narrative gaps. With Motion Books, some of this inferential work was displaced by prescriptive sonic cues.

For example, in *Captain Stone is Missing* (2012), the flagship series by co-founder Liam Sharp, when the protagonist enters a derelict building, the Motion Book tool automatically triggers a low, rumbling bass drone layered with intermittent metallic creaks, beginning as soon as the panel transition is completed. The intention is to convey menace and instability, but the cue operates prescriptively in the following ways: it frames the scene emotionally triggering fear and unease

FIGURE 4 Layered parallax in MadeFire™'s Motion Book VR demonstration. © MadeFire™, 2014.

before the reader can form their own interpretation from the art and text alone; being an obvious loop; it sets a temporal rhythm, effectively pacing the reading speed; and it reduces interpretive openness in the gutter, in that, instead of imagining what the 'silence' of the building might be, readers are supplied with a predetermined, cinematic atmosphere of threat across multiple panels.

Lessons from CyberComics and Motion Comics

From a narratological perspective we return to the idea that comics are multimodal narratives where sound is traditionally implied and not provided (Davis, 2023). The introduction of actual sound collapses the gap between diegetic suggestion (onomatopoeia, drawn sound effects) and extradiegetic control (the soundtrack). This collapse in *CyberComics* and motion comics likely left audiences who were used to traditional comics dissatisfied, because it redefined narrative voice: the comic is no longer narrated solely by the page and the reader's imagination, but by an external auditory track.

Although there is much criticism still in evidence on user forums regarding motion comics and MadeFire's Motion Books, perhaps the most damning condemnation of sound and motion enhanced digital comics was the shutdown of MadeFire in April of 2021, when a notice of insolvency was posted on their website. Despite distributing for major comic houses like DC, Marvel, Image and Dark Horse, having high profile investors and partnering with DeviantArt and MagicLeap, the company could not stay afloat. *The Beat*, a longstanding online hub for comics news reported that "in the end, readers have spoken with their wallets and eyeballs" and that 'motion comics' technology just never became a popular way to read comics" (MacDonald, 2021). Hague (2014) mused that this may have been due to the lack of clarity as to whether these are "comic[s] with a lot of animation or an animation with very little" (p. 76). Morton (2015) relates that "the contemporary motion comic rise and fall was fairly rapid, from sporadic releases (2001–2007) and a brief golden age (2008–2010) to a steady decline (2010–2014)" (p. 356).

A key lesson from these early experiments is how sound is best incorporated in the digital realm; in a manner that allows a more traditional reading experience. Sound is more seamlessly incorporated into the reading experience when it does not dominate or displace the visual-sequential logic of comics. Ellström (2010) suggests that successful intermedial hybrids allow the dominant medium to maintain control, while supplementary media enrich rather than overwrite. In practice, this means sound should indeed operate as a Deleuzian foreign body, as a framing element that reshapes perception without negating the visual temporality of comics.

Such an approach is further informed by Werner Wolf's (2015) idea of intermedial reference, that would suggest that soundtracks are more effective when they gesture toward the comic's visual rhythm (for example, ambient soundscapes aligned with panel transitions) rather than synchronised effects and literal voice acting. In other words, aurality in comics operates best when it is fluidly indexical and atmospheric, not dictating narrative, so that the reader retains control of pacing.

The role of sound in comics need not reflect the role of sound in film. As Chion (1994) points out, sound is not neutral: it shapes temporality, attention, and immersion. Unlike film, comics grant the reader absolute control over pacing, wherein "wherever your eyes are focused is the now" (McCloud, 1993, p. 104). If you use cinematic audio conventions, particularly dialogue or tightly synchronised effects, you risk removing that control and forcing a rhythm onto the reader.

By adding sound that parallels Chion's 'non-diegetic sound', the kind that contributes to mood and immersion without explicitly dictating timing, such as atmospheric sound (for instance, music beds and ambient and environmental

soundscapes), one better matches the agency of comic reading than moments of 'synchresis' that result with heavily event-based sound, thus creating the "spontaneous and irresistible weld" that results when "particular auditory phenomenon and visual phenomenon … occur at the same time" (Chion, 1994, p. 63). Atmospheric audio has the potential to act as what Chion calls "added value", a layer that enriches affect without disrupting autonomy (1994, p. 21).

The experimentation with rigid sound synchronization in DC's motion comics imposes Chion's "temporal dictatorship of sound" (1994, p. 13) onto a medium designed for readerly freedom in narrative construction. McCloud (1994) identifies the degree of "viewer participation" as of the highest importance and states that how comics address this "could play a crucial part of defining the role of comics in the new century" (p. 106). Perhaps this is why the longest standing consumer platform for digital comics is ComiXology—a platform that concentrated solely on silent, motionless print remediated editions.

Amazon purchased ComiXology in 2014, incorporated the platform into its store and it is still operational at the time of writing, and comic issues are still being sold in print. Whilst sound-enabled motion comics have largely failed in the market, a more open and accessible platform that favours user-agency in pacing (even though it mostly employs linear navigation), called Webtoons, is part of a recent market boom.

Webtoons: Sound integration in contemporary comics.

The foundation of Webtoons was laid down in early experiments with what came to be termed WebComics; independently produced digital comics posted online, usually by individual practitioners. Drawing on the scholarly works of Ernesto Priego, Josip Batinić, Sean Kleefeld and Leah Misemer, Benatti (2024) defines WebComics as "born-digital comics that are intended and designed by their authors to be read on digital screens such as computers, tablets or smartphones, and that are circulated through internet networks" (p. 6). With the earliest published versions appearing circa 1986 on the CompuServe network before the wider Internet existed (Kleefeld, 2020), WebComics, started as silent panel-based comics and early experiments with sound by their creators were only significant as digital comics first forays into navigation sounds, page turns and in providing the occasional limited low-quality sound effect. The data bandwidth available in this era likely limited WebComics engagement with sound.

Designed to be read on smartphones, the success of Webtoons, episodic digital comics that originated in South Korea, cannot be ignored in the discussion of sound integration. The platform, spearheaded by Naver's Webtoon and including companies like Kakao's Piccoma, and Webcomics, has successfully expanded into major markets such as Japan and the United States. In recent times, Amazon launched Fliptoon to the Japanese market and Apple launched a Webtoon page in Apple Books. Forbes Magazine has identified Webtoon as the world's largest digital platform with over 85 million monthly active users (MacDonald, 2023).

At present time, the Webtoon platform is experimenting with 'video episodes' which echo the synchronised dialog and sound effects of motion comics, but they are accompanied by a user survey to determine whether readers want to continue to engage with them, so it seems that their success is contested. These episodes often provide the option to 'play' the comic (the passive video/sound experience) or to 'read' them. The majority of Webtoons are silent and differ from traditional comics mostly in the single panel-based scrolling form of navigation.

However, there are also many Webtoon 'read' titles that employ musical soundtracks or environmental atmospheric

sounds, or a combination of both, such as in the *Soulwinder* series (2023). *Soulwinder* is an excellent example of the use of fluidly indexical and atmospheric sound, that complements the narrative instead of controlling it. I'll attempt to describe episode 35, *Gift*, to demonstrate why it is successful, but I recommend downloading the Webtoons app and experiencing the series first-hand to comprehend how sound operates within it.

The episode starts with a sweeping ambient musical drone that fades in as one scrolls into the first panel. It is accompanied by a bell-like instrument with a limited melody staying within a central musical key centre. Although the piece evolves, it does not distract reading, and it persists through the first few scenes and perpetuates long enough for even the slowest of readers. The overall effect is a floating otherworldly feel that supports the narrative

This drone continues until the reader reaches a scene where a seated conversation takes place on a veranda drowned in blue moonlight. After a few panels outdoor atmosphere fades in and despite the natural sounds of crickets and night birds

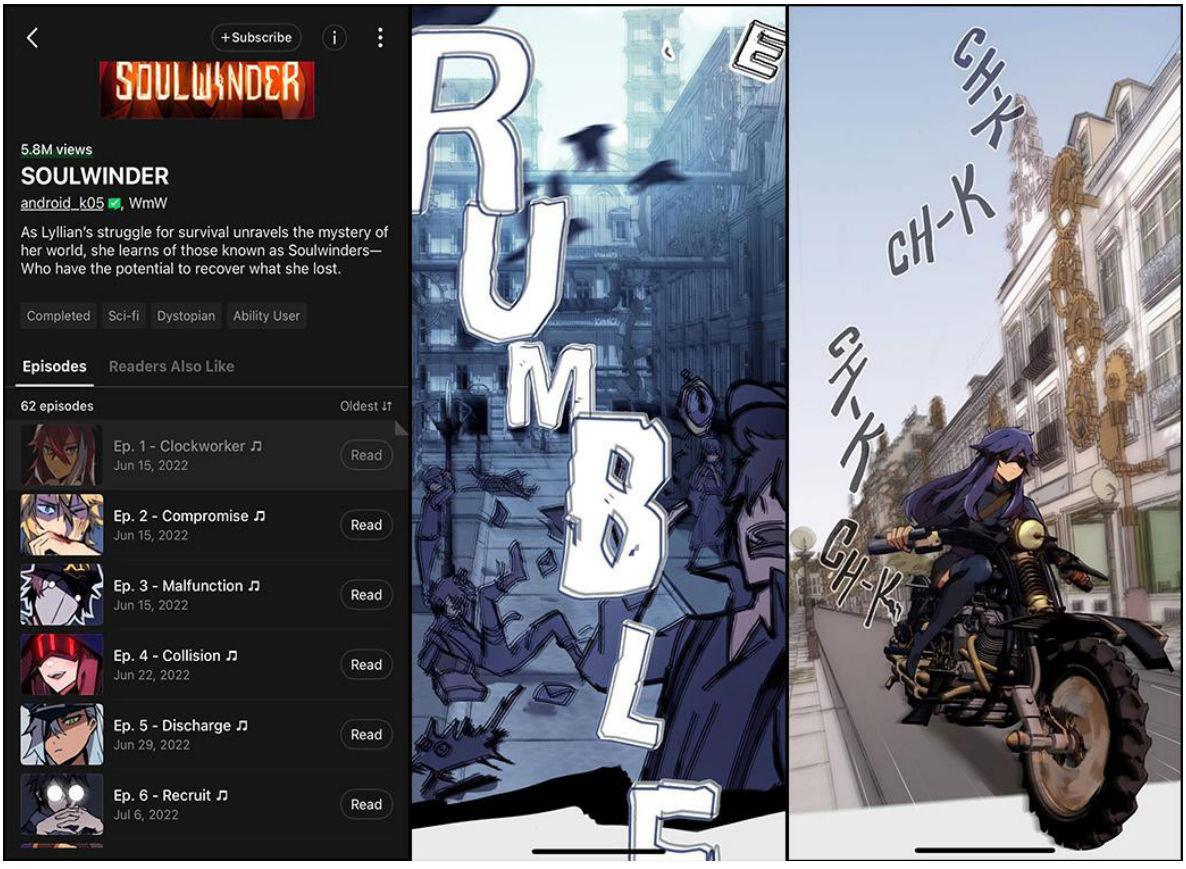

FIGURE 5 The *Soulwinder* original series, created by WmW and art by android_k05 2023 from the WEBTOONS mobile app. Naver Webtoon Ltd., 2023

there are low frequency traffic sounds that indicate the backyard scene is still located within a busy city. This is Chion's (1994) *acousmat*, a sound source out of sight, yet providing context to the narrative.

On the entrance of a third significant character, music is now layered with the atmosphere. This soundtrack is plucked strings and muted bells with a lightness that suggests the wonder and affection that is evident in a scene of reunited love interests. As the characters engage, the music slowly and unconsciously replaces the sounds of the night. When significant 'sound' moments occur, such as the tearing open of an important envelope, rather than a sound effect that might determine sync or that might be at odds with the mood, the imaginary onomatopoeic visual sound of a large "*RIIIP*" appears between panels.

As the next location arrives, panels that show the exterior of a dark mansion are accompanied by a silent reset that slowly lifts into a new atmospheric soundscape on entering this grim castle. Dank water drops accentuate a slightly industrial and ominous moving drone containing what sounds like descending glissando son filé on a cello. Muted clangs of large pipes ring, as low frequency distorted horn noises dive. This soundscape persists through the many panels of the scene and brings a heavy air of menace and tension to the confrontational exchange between characters, but at no time does it interfere with the speech-bubbled narrative.

The aurality in the slow-moving atmospheres and soundtracks used to enhance the mood of the visual narrative sits in perfect balance with the imagined visual sound effects. Webtoons such as *Soulwinder* point to the direction in which sound **in** comics can operate successfully to enhance the reader experience without stealing reader agency or completely transforming the silent tradition of comics.

The success of the Webtoon format is such that it has resulted in licenced music being used with Original Soundtracks (OSTs) often featuring popular K-Pop and K-Indie artists, as evidenced by any search on streaming music platform for the term 'Webtoon'. All sonically enhanced titles also provide a sound-off option, catering for those that wish to engage in silence, which also suits the on-the-go nature of the platform, for reading in crowded or noisy public spaces.

Having a platform where the consumer can choose to engage with synchronised sound, non-diegetic soundtrack or silent comics allows and encourages experimentation within the market without alienating the audience and is likely one of the ingredients of the success of the Webtoon.

Comic reading for the BLV community

Accessibility for the BLV (blind or low vision) community, who are unable to perceive the images, panel designs, layouts, and text, is another area in which technologies are advancing at the nexus of comics and sound. Mechanisms for generating text descriptions of comics for screen readers need to capture not only complex images, lettering, and mimetics, but they need to be able to use visual reasoning to convey the logic of page layouts and make decisions on the order and direction of narrative. To further complicate matters these systems must, at the same time, understand the closure-driven narratives conveyed by stylised artwork and dialogue (Iyyer et al., 2017).

Early attempts at using deep neural architectures to perform closure tasks based on narrative and character-centric aspects underperformed human baselines due to their inability to leverage context (Iyyer et al., 2017). The basic dataset for these experiments was called 'COMICS' and the performance of closure was greatly improved by focussing on improving the optical character recognition (OCR) capabilities. This

resulted in an improved dataset called 'COMICS TEXT+' that greatly assisted narrative understanding (Soyken et al., 2022).

More recent experiments with Multimodal Large Language Models (MLLMs) with capabilities in processing both natural language and visual information, have suggested that MLLMs are suited for the task (Rigaud et al. 2024). This is because of their conversational interface that can not only describe the comic, but clarify these descriptions via user interaction, such as "asking questions about their humour or meaning" (Ramprasad, 2023).

Rigaud et al. (2024) speculate that the enriched content descriptions resulting from MLLMs could further be used to create audiobooks and read-aloud eBooks with multiple assigned character voices, captions, and sound effects. It is worth mentioning the alternative haptic tactile renditions of comics for the BLV community, which already exist. The potential and the restrictions of sequential pictorial storytelling in such accessible works is an area of research for the Berlin-based Jakob Dittmar (2019).

Sound in AR comics

Augmented Reality (AR) comics also offer new possibilities for sound, although whether they fit the formalist definitions of 'comics' is highly contestable (Cohn, 2005). Formalist definitions of comics emphasize evaluative criteria, such as sequence, closure, and dependency on stillness and silence that AR adaptations may violate (Meskin, 2025). The physicality of the technologies that they depend upon for the AR layer, such as viewing pages through a smartphone or holding the comic in front of a webcam, also "rips the reader out of the reading experience" interrupting immersion (Helms, 2016, p. 61).

However, some exponents of new comic technologies, refute the requirement of equating their experience with that of their printed ancestors and provide their own caveats that determine how to best design digital narratives. The *Digital Comics Manifesto* (2016) by Ezra Daniels (the designer of *Screendiver*-the digital comics directory) co-signed by 215 creators, states that a digital comic should "transcend print" looking to "stake out new territories" using the full features of the chosen platform, but still relates that it should "never take temporal control from the reader" as that is a defining aspect of a digital comic. In the accompanying video, he states that "passive video, voice acting and literal sound effects yank temporal control from the reader, making the work something different than comics".

When Daniels (2016) also relates that "Print-formatted page layouts and skeuomorphic folding-paper transitions have no place in the digital realm" it is exactly those things that apps like the Veve AR comic reader seek to accommodate. Such apps allow one to place a digital comic on a flat surface viewed through a smart device screen and read it via page turning gestures. There is no sound involved in the Veve reader, however an AR/VR reader called Livro transforms a print remediated digital comic into a physical object in space, replete with page turn sounds, such sounds being identified by Hague (2014) as part of the performance and physicality of comics, and is a shining example of McCloud's idea of how sound can "sneak in through the back door" (2000, p. 229) via navigation.

A different level of AR was leveraged for the comic *Retinex* (2010) from Metaverse One, with scannable image codes on the page linked to story-relevant animations including sound and interactive 3D models. The dystopian *Retinex* was arguably the first comic to use AR for story-integrated elements. This print comic-book adjacent style of AR did not interrupt reading in a traditional manner by dictating pace with sound or animation and the printed component could be read without the technology.

Although I located the book, I was unable to locate the custom-app developed for the AR layer. As with many

digital multimodal experiments with comics, many AR-based projects are reliant on proprietary technologies, their functionality disappearing as the apps that drive them become defunct. We must now rely on reviews and video

FIGURE 6 Motion graphics and sound, enabled by scanning a MarvelAR title with a personal device. © Marvel, 2012.

records of such technologies to get a sense of their efficacy and operation until there is a move towards 'critical librarianship' and collaborative practices in digital comic preservation (Gebhart, 2019).

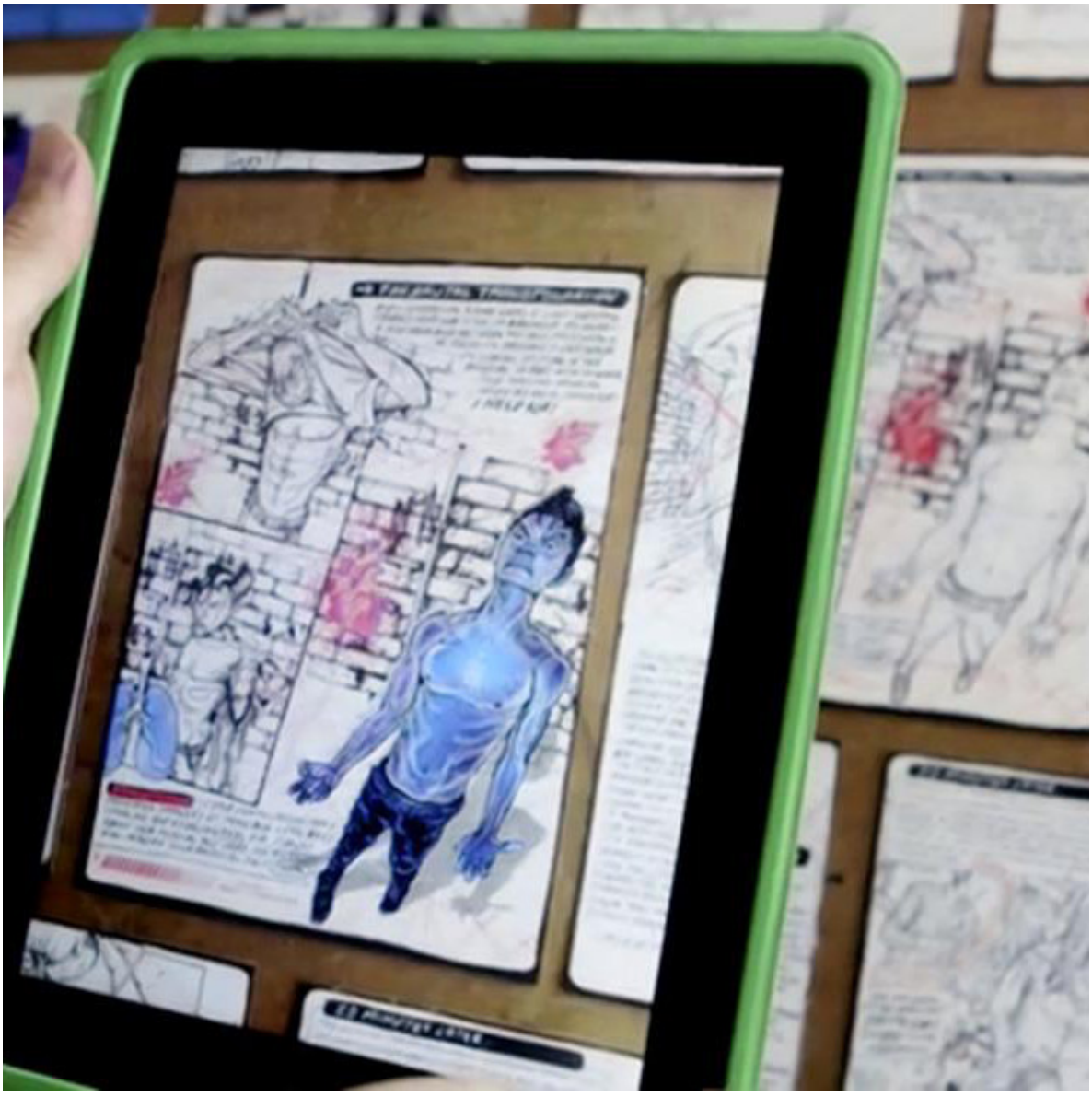

FIGURE 7 Image scanned AR in *Modern Polaxis* activates sound, colour and motion. © Sutu, 2014.

One such defunct technology is the Marvel AR app created in conjunction with Aurasma, announced in 2012 (Constine), and shut down and removed from stores in 2015. It was the first AR app launched by a comic book company. Embedded video series recaps and introductions, character bios, concept art and creator commentaries and interviews could be experienced by pointing the personal device camera at any page with the MarvelAR logo. The content was reminiscent of DVD bonus features and was always extraneous and independent of the reading experience, although sometimes a panel of a comic would be overlaid with an animation and sound, bringing the page to life.

Gray (2010) speaks of how paratextual sound and video elements such as these legitimately extend and shape a primary work by prefacing, contextualising, and extending the experience of the central text. His work frames them as legitimate components of a text's reception ecology, claiming "paratexts surround texts, audiences, and industry, as organic and naturally occurring a part of our mediated environment" (p. 23).

However, the print-based AR comic can also be used creatively, to invent new forms of dual-layer narrative, as with *Modern Polaxis*, a comic about a 'paranoid time traveller'. The printed text serves as the character's private journal, but his 'paranoid delusions and conspiracy theories' are accessible via the AR layer, providing two parallel readings of the text. As the sound dimension is only accessible when accessing the media-rich AR layer, the 'delusional' aspects of the character become richer and carry more weight, much in the same way that the world of sound becomes richer and more complex in its considered perception. This reading of heightened senses is accentuated by the print narrative being silent, motionless, and black and white, while the delusional AR state is sonically rich, animated and in colour. This dual narrative speaks to the ability of sound to reveal or externalise what silence leaves hidden. The AR layer of sound is cleverly used to represent interiority, hidden psychological worlds and subjective states which remain invisible in the purely visual register (Altman, 1992; Gorbman, 1987; Lastra, 2000). The sounds used are not obviously linked to the visible world of the print comic, recalling Chion's assertion that "the suppression of ambient sounds can create the sense that we are entering into the mind of a character absorbed by her or his personal story" (1994, p. 89).

Modern Polaxis is one of many AR comics that uses image-recognition for augmentation triggers rather than QR codes, which enables a more immersive experience without visual interruptions when launching AR features. Whilst it relied on an iOS app custom-coded by the comics creator Sutu (aka Stuart Campbell), many AR comic titles were reliant on third-party apps by companies such as BlippAR for functionality, creating a fragile state for such digital media artefacts.

Creators of sonically augmented comics need to consider best practices for publishing and preserving their work by considering what Margret Hedstrom refers to as 'temporal interoperability', the "ability of current systems or legacy systems to interoperate with future systems that may use new formats, data models, languages, communication protocols, and hardware" (as cited in Sabharwal, 2015, p. 15). Sabharwal also mentions considering the dangers of "commercial actions such as mergers, failures, and other compatibility issues affecting accessibility to legacy files" (2015, p. 15).

BlippAR went into administration in 2019 and sold its intellectual property assets to an investment firm, and as a result, the app that powers their AR comics (such as Ram Devineni's *Priya's Shakti* and Daniel Corey's *BLOODWORTH*), is no longer available for the public to download leaving these titles in limbo. Sutu went on to become director of an AR/XR company called EyeJack that now provides the infrastructure for the functionality of all his comics, artworks, and installations, also having created VR experiences for companies such as Marvel and Disney.

Augmented comics have also been created inside virtual reality spaces, for example inside VRChat. VRChat is a social VR space where users create environments, avatars, and new experiences. Within this platform creators are experimenting with 'comics-as-installations', panels and sequential imagery arranged in 3D space, often with interactive triggers. Although it is a VR space, these 'comics' often use augmented features (sound overlays, pop-up effects, or interactive speech bubbles) that sit as an AR layer on top of static comic art.

Typically, in the AR/ChatVR environment sound exists as environmental soundscapes, indexical cues, and voiceover. The atmospheric soundscapes are typically spatial audio employed to create mood; for instance, if you move your avatar into a panel depicting a forest, you might hear layered bird calls and rustling leaves. Some AR comics experiment with live or pre-recorded voiceover, merging comics with aspects of performance art. Indexical sound cues are responsive to choice-driven triggers. Footsteps, door creaks, or puddle splashes might be triggered when a 'reader' walks into a panel environment.

It is important to note that these sounds operate in a spatial sound environment, which does enhance the immersive experience. Ambient sound affects the listener's overall perception of atmosphere (Pancirolli et al., 2023, p. 321). Spatial audio engines used in platforms like VRChat make sound both directional and locational, "stimulating the subject's senses in a realistic and expected way", deepening the atmospheric effect (Gilles, 2020, p. 207).

Another developing sound technology that is possible to utilise with AR comic cues is called *Ultrahaptics*. This is feedback derived from sound pulses to simulate the feeling of touch including the feel of dials and buttons (Fink, 2018). In terms of user experience (UX), sound can also be useful as input for voice command "as well as inferring the user's context" (Ritsos, et al., 2011, p. 3).

VR comics and spatial sound

After partnering with AR platform company MagicLeap, MadeFire started creating AR comics, moving their motion comics platform from mobile devices across to VR headsets. They also migrated their use of synchronised sound events in 'motion books' to AR. However, there was restraint that developed from an awareness of the importance of closure related to traditional comic reading that limited the amount of sound they included, including the altogether avoidance of voiceover. In an interview with Ben Wolstenholme, founder at Madefire, recognising the importance of the concept of closure he related "we quickly realised that we wanted everything to be reading, not watching … once you start getting into voices … it becomes passive … you lose the engagement with the reader" (Cecchini, 2012a).

VR/AR comics could be considered a radical rethinking of the gutter between comic panels as physical space. The readers eyes no longer simply move from one panel to the next or across the flip of pages. Instead, the readers walk through panels in a new form of embodied immersion in which sound cues play a key role. The space between panels or scenes becomes a three-dimensional, navigable zone that extends beyond its conventional function as a marker of closure. Dittmer (2010) described the gutter as the geographical space between panels where readers actively construct meaning, while immersive environments transform this liminal space into an auditory bridge, rather than a silent void. In an VR/AR comic adaptation, sound design can occupy gutters as spatialized audio cues, shifting from indexical signals to atmospheric ambiences that guide users through narrative transitions.

Groensteen (2007) has argued that the 'spatio-topical system' of comics organizes meaning across both panels and their intervals. In VR/AR comics the gutter in can function as a sonic threshold, dynamically situating the reader between

narrative worlds and opening possibilities for fluidly indexical and atmospheric soundscapes that are not only narratively functional, but also immersive in ways traditional comics cannot achieve, as spatialized sound provides orientation and continuity in the perceived three-dimensional space and can enhance presence and embodiment (Langiulli et al., 2023; Wincott et al, 2021). This includes the use of spatialised music cues.

Visual representations of music and soundtrack for traditional print comics

Hague (2014) examines works of sound **with** comics in which the external soundtrack is a pre-paced reading of the com- ic books textual elements including dialogue, captions and narration, name-checking the Marvel/Golden Records Marvel Age Comic Spectaculars series, an evolution of the read-along children's books from 1960s that were paired with 7-inch 33 1/3 RPM vinyl records. In such books Hague highlights the new issues that they introduce such as the impact of the tone and character of voice acting and the authorial intru- sion that occurs with the breaking of the fourth wall, directly addressing readers, just as Stan Lee loved to do in the print comics he edited. This issue of the intrusion of the voice is also, by nature, shared with the audio comic (sound **as** com- ic). Hague further delineates between such comics that re- produce the visual language of the work faithfully and those that don't, noting the possibility that soundtrack can provide information not seen in the visual work. One could consid- er this a new take on the idea of McCloud's 'infinite canvas'. Such works, operating in a loose but connected synchronisa- tion with a rendering of the textual narrative are prescriptive and set a rigid tempo that favours the creators' preferences, rather than the readers. This is not the case if the sound **with** comics is instead a purely musical soundtrack.

Introducing music soundtrack to a traditional print com- ic-reading experience is a simple way of adding new dimensions to the work, without having to completely change the paradigm. A comic can be enhanced by a music soundtrack without the need for accurate sync or dictating the pace of narrative within the reading activity, with some caveats.

Whilst the rhythms and regimens of art and literature in comics become superimposed on each other (Eisner, 1990), in a similar manner, aspects of music, with all its own inherent complexities, can and will be superimposed on the combination of art and literature of comics when soundtrack is introduced.

Regardless of individual reading preferences, comics have their own visual rhythm and musicality. Many authors speak to this quality including Groensteen (1999/2007; 2013), and Hatfield (2005), who examines rhythm and sequence and analyses the ways alternative comics rely on repetition, variation, and thematic recurrences; whilst Summers (2015) examines ways in which music can be written into the narrative by using the visual iconography of music in comics, as does Pellitteri (2019), who ventures "synaesthesia, in the particular case of music in comics, necessarily implies the displaying of music symbols, song lyrics, music scores, and/or narrative situations in which characters are clearly engaged in musical activity" (p. 528). The way in which the musical rhythm, and even the sonic character of song can be translated into the visual rhythm of a comic has been exemplified by Horton and Gray (2022) in their analysis of Alan Moore's comic remediated rendition of a 1967 song by British group The Move, titled "I can hear the grass grow".

Writer and artist Dave Chisholm describes representing music in comics as 'fundamentally impossible' and speaks of instead creating 'points of synchronisation' between the temporality of music and the temporal aspects of a still drawing, or alternately, capturing it's mood by other means or by the use of musical notation (as cited in Holley, 2020).

Chisholm's *Chasin' the Bird* (2020) is a prime example of the synaesthetic representation mentioned by Pellitteri. In this graphic novel Chisholm depicts sound not just through conventional notation or onomatopoeia alone, but by constructing musical meaning through changes in layout, colour, and drawing styles that correspond to different narrators and musical perspectives. Each chapter recounts Charlie Parker's time in California from the viewpoint of a different witness, and each perspective is rendered in a distinct visual language.

Viewed within Groensteen's spatio-topical framework the organisation of the page changes according to the subjectivity of the narrator, meaning that the reader's perception of Parker's music is mediated through the spatial configuration of the comic itself. The site of panels, their density, and their visual intensity function as equivalents for tempo, timbre, and emotional tone. Because the spatio-topical system allows panels to interact across the page as a network rather than a simple sequence, the reader experiences these stylistic differences simultaneously as narrative perspective and as sensory variation.

Because these forms occupy space within the panel and sometimes extend beyond it, they alter the spatio-topical structure of the page itself. Panels containing music often become visually unstable, with backgrounds simplified or removed, allowing the abstract shapes to dominate the composition; the spatial hierarchy of the page shifts so that music becomes the primary organising element. Whilst the reader cannot literally hear Parker's improvisation, the spatial disruption of the page produces a synaesthetic effect in which visual intensity can be interpreted as sonic intensity.

Reorganisation of space evokes rhythm, volume, and improvisation. When Parker's solos are depicted as expanding colour fields, when panel borders dissolve, or when styles shift between chapters, the spatio-topical system itself becomes the medium through which sound is imagined.

In this sense, *Chasing the Bird* demonstrates that the synaesthetic potential of comics lies not only in symbolic imagery but in the structural properties of the page. By manipulating panel form, site, and spatial relations, Chisholm creates a visual rhythm that the reader interprets as musical rhythm, reaffirming Groensteen's claim that comics meaning emerges from the interplay of space, placement, and relational structure rather than from linear narration alone.

To Eisner (1990), tempo and rhythm are important aspects of comics and comics reading and he relates how "timing and rhythm are interlocked" (p. 30). He speaks of the rhythm of story and of panels and how they can be altered by perspective or by visual shape; he speaks of the rhythm of a total page; he describes beats, pauses and metre in comics, of staccato vs. legato and of how timing is important in conveying a "specific message and emotion" (p. 34).

With Eisner identifying the multiplicity of factors affecting the rhythm and pace of reading, it is no wonder then, that comics are highly interpretive manuscripts. Groensteen (2013) concurs that the rhythm of comics is complex matter, also stating that "the discontinuity that is the basis of the language of comics ensures that rhythm is a central element of its discursive resources" (p. 193). Adding to the complexity of construction, rhythm and musicality can also be created utilising the words of the narrative, as related in Miodrag's detailed analysis of the work of George Herriman (2013).

Rhythm in the visual elements of a comic is also performatively constructed. Eisner (1990) provides the understanding that reading comics is a perceptual activity, also pointing out a parallel in the reading of musical notes as he highlights the interpretative nature of both to the reader, explaining that "to the different formats configured by the artist there correspond different rhythms of apprehending the material" (pp. 26, 154–5). Summers (2015) examines participatory ways of incorporating music including the inclusion of score for those

who can sight read, or the imaginative reader recall of popular music references embedded in the narrative.

Mee & Robinson (2024) speak of how "the sound affective event is also categorized by the viewer's cognitive encounter

FIGURE 8 Singer/lyricist Claudio Sanchez of music group Coheed & Cambria created a series of music albums and co-authored a comic series transmedia experience. © Equal Vision Records/Boom Studios, 2002.

with it" (p. 4). When we introduce the dynamic of reader reception, pace, and interpretation into the multimodal three-way interplay between text, artwork and layout and music, it would be fair to say that every reader will achieve a different (if not totally individual) and unique experience.

Given this intricate web of engagement, a suggested piece of music does not have to be explicitly composed to match a graphic narrative as soundtrack. Much like comics themselves the way music is received is very much up to the listener as "in transmedia storytelling, as in music, it is what [the audience] brings to the work, the new meanings they find, the embellishment and deeper layers they'll peel" that will inform the experience (Weaver, 2013, p. 44).

That being said, there is a tradition of creators creating thematic works that relate to the subject matter of a comic book that date back to the 1980s, for example, comic book writer and musician Timothy Truman supplied an issue of his own comic creation, *Scout*, with a two song phonograph flexi-disk in the pages of the issue by his band the Dixie Pistols, following it up with a vinyl LP called *Marauders* (1987) based on themes in the comic and containing an eight page mini-comic that detailed events post the comic's first series. There is no pretence of synchronisation between the works outside of themes in the songs that relate to scenes or chapters in the story and the pace of both reading and listening are under the control of the consumer of the works.

Conversely, comics can also be the vehicle to convey a storyline first established within a music project, as is the case with the progressive rock band Coheed and Cambria. Their primary singer and lyricist Claudio Sanchez co-authored the science-fiction comic series *The Amory Wars,* to add depth to the narrative established in the lyrics and musical motifs their musical works, with each of a series of five studio albums correlating to a to a chapter in the saga. The debut LP *The Second Stage Turbine Blade* was released in 2002, with the comic book series following in 2004. Together the comics and music provide a transmedia experience that can be experienced simultaneously.

One novel approach to creating a comic soundtrack is to encourage the reader to build their own soundtrack using references to popular music interwoven with the narrative, which is exactly what *Radio Apocalypse* by Ram V (2021) seeks to do. Inside the cover the author relates how the book was written whilst listening to music and asks:

> Let's make a deal. I'll tell you the best story I can about lives, loves, loss, and joy. And you—when you find an image or a line referencing a song or a character playing music on their boom-box, walk-man, or hand-held radio, I want you to find that song wherever you listen to music, and I want you to play it. That's how this story of the last radio station standing silhouetted at the far horizon is meant to be read. Plugged in and turned up.

Whilst Ram V uses cues within the comic book to suggest the soundtrack to the user, the technologies of QR code linking and the ease of creation of playlists on music streaming platforms, particularly Spotify and YouTube Music have been utilised in other comic titles to provide instant access to anyone with access to a capable portable device.

In issue five of *I Hate Fairyland* (Young, 2023) readers are told to scan a QR code and "rock out as you follow along". The code links to a landing page with choices of streaming services to listen to "This World" by Jack the Radio, a song that was written for the comic in place of dialog in the issue (in which there are no speech balloons), and successfully echoes the unbound spirit and tone of the storyline and characters with simple lyrics and anthemic psychedelic riff-rock. David Aja created soundtrack playlists of existing popular music tracks for many of his co-created *Hawkeye*

issues in the series that ran 2012 to 2015. These playlists were also compiled in his Omnibus collection (2023) and then uploaded by fans on Spotify and YouTube in acts of cooperative fandom.

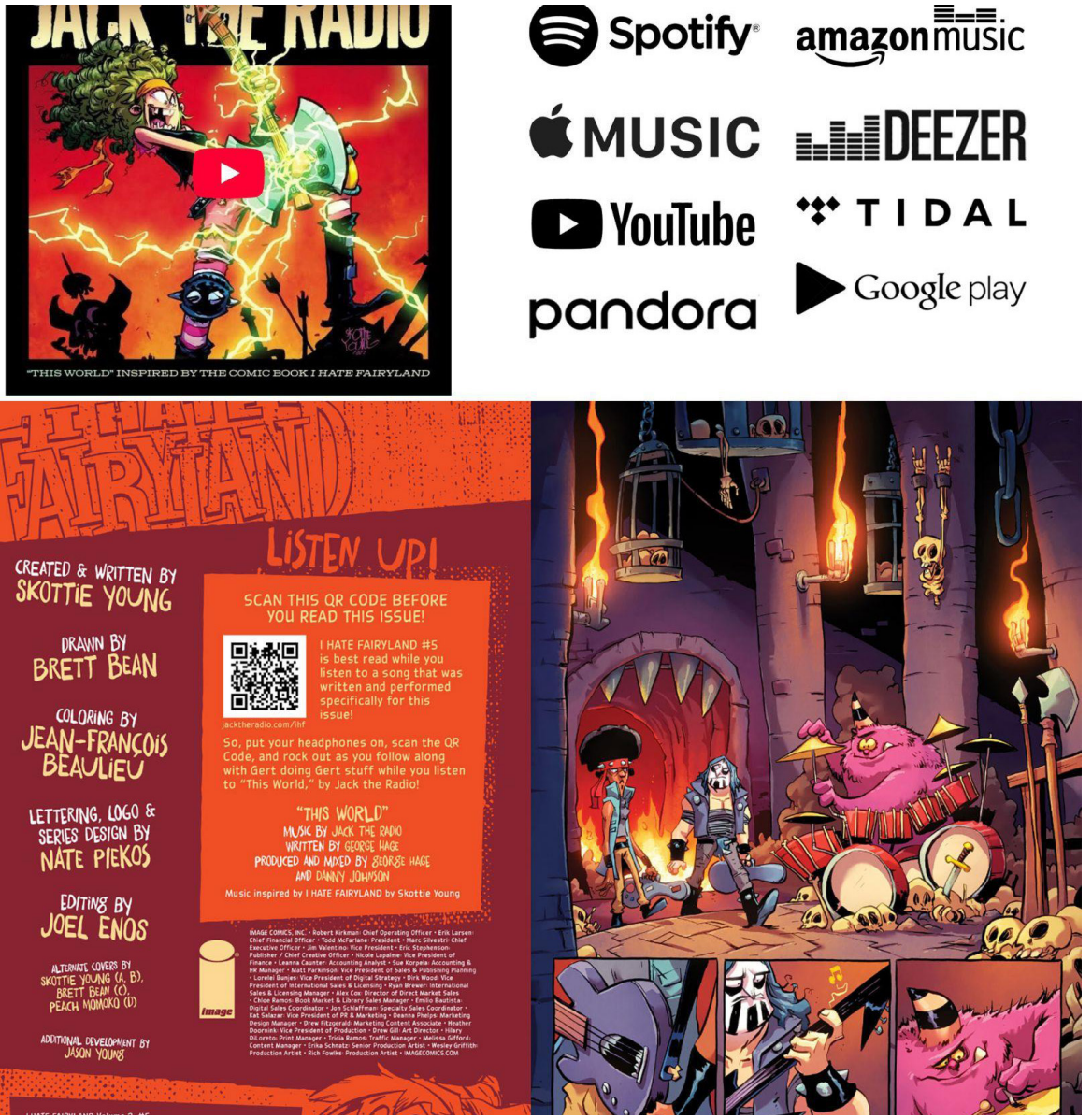

FIGURE 9 Vol 2, Ep.5 of *I Hate Fairyland* by Skottie Young uses a QR code link to a dedicated soundtrack by the band Jack the Radio. © Image comics, 2023.

A recent development surrounding the 'suggested playlist' trend is that fan afficionados are creating and publishing their own playlists relating to comic titles on streaming and social media platforms. A prime example of this is the Parker Records YouTube channel (@parkerecords, n.d.) dedicated to this pursuit. Influenced by the affective tastes of the person creating the playlist, the texts are taken beyond the author's agency into a realm of participatory sonic remediation. The ability of fans to transform a text in such a way is reminiscent of jazz musician Ornette Coleman's (1983) concept of 'Harmolodic Democracy' in which "all ideas have lead resolutions… each player can choose any of the connections from the composers work for their personal expression" (p. 23).

The phenomena of soundtrack in comics as a participatory medium echoes the sentiments of Henry Jenkins who observes that "fandom does not preserve a radical separation between readers and writers". One could also argue that the audience is not transforming the experience of media consumption into the production of new texts in remediation but instead creating shared multimodal experiences.

The very personal and subjective nature of comic soundtrack experiences, mean that not all readers will be receptive to them, and indeed, given the failure of digitally enhanced comics to find widespread market acceptance, it is also a fair assumption that many readers would prefer to experience comic books in their original form, both visual and silent.

In the purposeful coupling of external sound sources with comics there is always a struggle taking place to determine the dominant medium. In a 2011 presentation Hague noted that "more consciously deployed sounds, such as those that are found in comics with attached records or CDs, can… serve to modify the nature of the reader-text interaction and renegotiate relationships of power" and highlighted that it was an act of "double reading". This is recognised by Pellitteri (2019), after examining the potentiality of the visual sound mechanisms representing radio playback in the Italian comic *Lazarus Led*. Pelletteri mentions the authors note at the beginning of the issue suggesting that the reader listen to a U2 album whilst enjoying the story, calling this double reading "a crossing between intradiegetic and extradiegetic music, so as to highly favour auditory enjoyment".

The totally sonic comic: spatial enhancement and sonic closure

Besides the BLV community, it follows that it must be a niche or highly differentiated market that supports the premise of a comic book that is entirely adapted into the medium of sound; and yet, such adaptations exist in the form of the 'audio comic'. A purely sonic 'comic' presents the most radical challenge to our understandings of the traditional medium. It could be easily argued that this is, in fact, a total departure from the traditional medium, presenting a sensory shift; a new form of closure using narrative information that is heard but never seen, causing Hague (2014) to relate that he was "yet to encounter any works that could truly be said to display the formal properties of comics through sound" (p. 82). Indeed, there are some who exclude the audio comic even from the definition of digital comic (Aggleton, 2018).

The aesthetics of audio comics hinge on the sonification of spatial form. In visual comics, spatial juxtaposition organizes meaning, using panels, page layouts, and gutters to provide structure (Groensteen, 2011). In the imposed linear narrative of the audio comic, this spatiality can be at least partially restored through spatial audio techniques: sound localization, depth of field, and environmental reverberation.

Research on enhanced audio description suggests that spatial audio can offer an immersive proxy for visual space, enabling listeners to perceive layout and proximity, as well as character and object motion through sound (López et al., 2021). When applied creatively, these techniques become

not merely accessibility tools but artistic strategies of narration. Hague (2014) reminds us that the audio comic is not mono-sensory, as "sounds are spatial in character as well as temporal" (p. 81) and this spatial character is what can be enhanced with surround sound technologies, such as Dolby Atmos™.

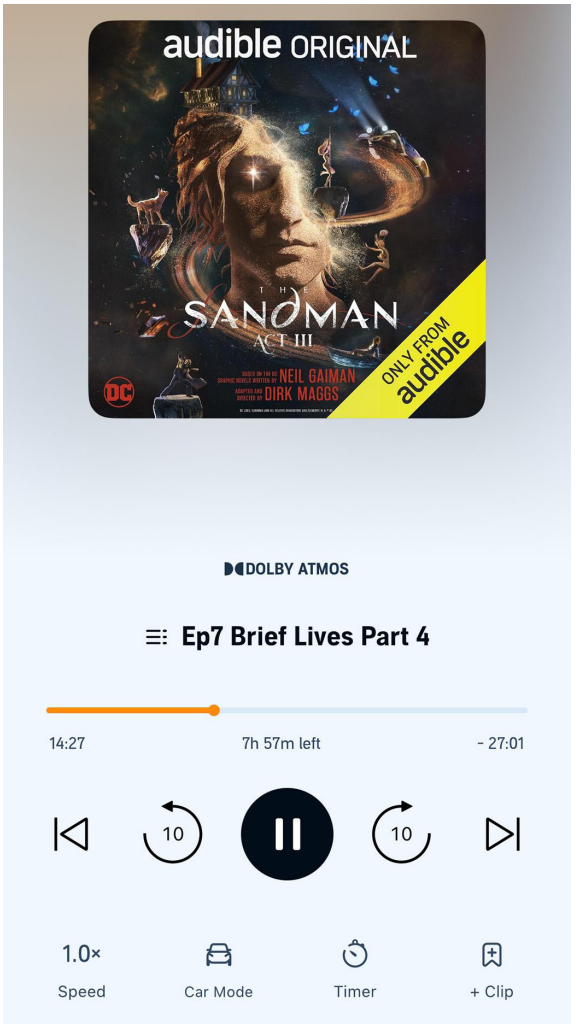

FIGURE 10 Act III of Neil Gaiman's DC Comics series *The Sandman* was remediated to an 'audio comic' rendered in binaural Dolby Atmos. Image © Audible Australia, 2022.

The binaural illusion of Dolby Atmos™ 'immersive stereo' is partially convincing with carefully positioned loudspeaker systems but when experienced with Dolby Atmos™ capable headphones the listener experiences realistic 'surround sound'.

In audio-comics and audiobooks, the audio data is carried in a stereo Dolby AC-4 bitstream with appropriate stereo content plus the metadata that a decoder uses to translate the virtualized experience. Technical delivery method aside, the net result is a 'real world' sound experience, with the listener placed 'inside' the sound environment.

Using the immersive stereo of Dolby Atmos™, titles such as Audible's *The Sandman* (Gaiman & Maggs, 2020) and *The Boys: Deeper and deeper* (Burns, 2022) demonstrate how spatial audio, voice casting, and soundscapes can constitute a sonic equivalent of the sequential art of their print comic origin. Characters are distinguished by timbre and shape, while environmental shifts replace panel borders, guiding the listener across "aural gutters". Movement is recreated not by the visual perspective of sequential panels and the reader closure that occurs in the space between them, but by 'phantom panning', the positioning of the point source of a sound object in real time, in three-dimensional space, by manipulation of head-related transfer functions (Dolby Laboratories, 2018).

This radical remediation of comics into sound again raises the ontological question of whether 'audio comics' can be still be considered comics. Intermediality theory provides a framework, suggesting that such works exist in a 'border zone' where medium-specific traits are refashioned but not erased (Rajewsky, 2005). Rather than treating the absence of visuals as a loss, audio comics can be understood as remediated hybrids, in which comics' narrative logic— its rhythm, framing, and juxtaposition—can survive and even thrive in a purely audible register. Spatial audio technologies go some way to demonstrate that the essence of comics may lie not

in their reliance on images, but in their structuring of narrative through sequential rhythm and spatial organization; qualities that can be re-imagined through sound.

Such binaurally encoded renderings of the conventions from traditional print comics help to convey the designed imaginings of artists, colourists, panel-designers and letterers, providing the potential for more faithful remediations of print titles for the BLV community and other fans of sonic comics. Comic titles created purely with sound also reveal new opportunities for imaginative closure.

Replacing visual and tactile analogues with sound requires participatory reconstruction grounded in temporal experience, inference, and imaginative embodiment. It demands more interpretive work, as the listener must reconstruct not only what happens between sounds, but must explore spatial, visual, and emotional meaning, expanding the domain of closure beyond visual gaps toward multi-sensory construction.

Unlike the static simultaneity of comic panels, sound unfolds linearly, and I have repeatedly highlighted how this can conflict with the temporal reader agency of the print comics experience. Sound that is fluidly indexical and atmospheric work better in most hybrid comic forms.

However, the audio comic highlights new heightened experiences possible with sound. Here, Jonathan Sterne's (2012) notion of the 'sonic imagination' and imaginative whole-world construction come into play, with the listener inferring continuity, not by visual scanning, but by anticipating and retrospectively interpreting sonic cues.

Furthermore, the technologies of surround and binaural audio embody Walter Ong's (1982) assertion that "sound incorporates, sight isolates" (p. 70), as they assist the listener to entirely construct the visual, map the spatial, and interpret the unseen. A well-designed and rendered audio comic can travel beyond simple remediation, transforming the listener into an active composer of story worlds.

Conclusions

I would venture that most readers already enjoy engaging with the sound **of** comics, the sensory wonder of the physicality of comics in print. However, sound **in** comics and sound **with** comics provide new sensory possibilities that can leverage and embrace the talents of composers, musicians, sound designers, audio engineers, music producers and programmers and provide new layers of participative narrative design.

The biggest obstacle to sound **in** comics is in disrupting or destroying the imagined sonic world of traditional comics, and early experiments such as Marvel's CyberComics and DC's Motion Comics, whilst they broke new ground, demonstrated the pitfalls of using sound as 'add-on', jarring with the visual rhythm and interrupting reader agency in temporality with linear sound and attempts at synchronized sound effects and dialogue.

The auralised version of the Webtoon such as *Sidewinder* provides a prime example of how non-indexical soundtrack and atmospheric sound can work in harmony with visual sound conventions, a meeting of imaginary and actual sound, that enhances the visual comic tradition without replacing it. It allows a degree of reader agency in navigation and pacing and provides a different kind of closure, allowing the creator some degree of authorial control in the narrative experience via thoughtful sound design and composition.

Sound **with** comics, such as reading print comics with suggested music playlists is a reader-mediated experience; one can opt in to experience it or abandon it for traditional silent reading. Having tried it on many occasions, I can affirm that at times I felt that it did enhance my reading experience, when the music didn't detract with the visual and textual narrative

in print. For me, this was typically slower paced instrumental music with controlled dynamics and a less prominent mid-range frequency range and tending towards the ambient. I found that, at times, it could also detract from the reading experience.

Sound **as** comic—the totally 'sonic comic' remediated entirely as audio—enters a different realm altogether, with a new kind of limited-sensory closure in which the imagination draws solely on sound, allowing the reader/listener to close off the visual world entirely if they wish, and to co-construct a world using a different set of mental parameters. It is in this sonic world that the immersive properties of surround sound, can push past the old boundaries of the stereophonic. It can wrap the listener in convincing three-dimensional environments that echo real world experiences with sound, and can greatly enhance the active participation in narrative that comes from such imaginative closure.

Addendum: Sites for further potential exploration—auto-generated soundscapes for comics

Referring to my assertion that sound works best as companion to comics when it is fluidly indexical and atmospheric, the possibilities of procedurally generated adaptive soundscapes using atmospheric and indexical sounds could be explored. Technologies to drive this would also sit at the intersection of computer vision, MLLMs and accessible computing. The basis of this idea already exists in Webtoons where the soundtrack shifts depending on what panel, page or screen a reader is on, providing composed ambient layers that enhance immersion without dictating precise timing.

This process could take place without the sound specifically being author-designed for the scene with contextual multimodal analysis of whole images including composition, mood, environment, and narrative context, resulting in a textual description. This text caption could then be fed to a trained MLLM to contextualise and render that description acoustically via synthesis or a limited recombinant sample library.

Methods for achieving a whole scene analysis are being explored for photographic images, such as using 'bottom-up attention combined with 'top-down' attention based on deep neural networks. The former is the automated visual feed-forward cataloguing of "unexpected, novel or salient" objects in the scene, whilst 'top down' refers to searching for the predictive non-visual or task-specific context, that is used to weight each identified feature in creating the image caption (Anderson et al., 2018). This could be further informed by Cohn's (2021) work evaluating characteristics of visual narrative processing and proficiency and his exploration of factors of cultural variation.

Krause et al (2017) have developed a model that pulls apart both images and paragraphs searching for semantic regions in images and uses a neural network to reason about language by scanning multiple levels or layers to determine dependencies. The intent for the output of this model is identifying images with long, descriptive passages, possibly resulting in a richer and more detailed, contextual language guide that could be used for the creation of corresponding soundscapes for the image.

At present time these systems are being tested on photographic images. As to whether any of these systems could successfully decode the abstractions of comic books, given the heavily stylised renderings of artists and colourists, remains to be seen, and comic page descriptions would have to be informed and supplemented by MLLMs performing panel-analysis, advanced OCR and closure tasks, as discussed earlier.

Once a valid textual description of a comic book panel, sequence or scene has been created, the question of how it will be recreated with appropriate added indexical soundscapes arises. The answer is in emergent text-to-sound AI generation tools, such as Meta's open source AudioCraft, declare-lab's TangoFlux and other AI audio tools from Stability AI, OpenAI, Google DeepMind, and ElevenLabs, Although the generative models and methods differ, such tools could be incorporated into the back end of a contextual soundscape generator for comics.

The integration of automatically generated sound into comics has the potential to enrich the reader experience but could also have the potential to interrupt or even derail it. One major concern is the indexical quality of sound. Automatic systems frequently default to literalist mappings which can narrow or flatten interpretive possibilities. Literal mappings of sound risks being overly didactic, guiding readers towards a single interpretation. Such mappings could override or contradict a creator's use of the narrative capabilities of comic's 'silent sound' conventions, and likewise, the creator's ability to exploit Chion's (1994) 'added value', in which sound "interprets the meaning of the image and makes us see in the image what we would not otherwise see, or would see differently" (p. 34).

Another possible pitfall of such automatically generated sound involves the granularity of automation. Autogeneration of sounds cannot just work at the panel level, as panels are not discrete temporal units; they are part of a larger sequential system where the meaning emerges in the gaps and gutters between panels as much as the image itself, with the gutters becoming "the site of semantic articulation" (Groensteen, 2007, pp. 168–169). If sound is generated per-panel without regard for broader narrative rhythm, the result may be incoherent or overly fragmented, undermining the way in which the audience constructs their personal 'virtual soundtrack' (Cook, 1998).

Such a system may also create output that is an aesthetic mismatch with the text. Automatically generated sounds may fail to capture the oeuvre or stylistic tone of a given comic (for example ironic distance, or surrealism), leading to incongruity between the visual and auditory layers. Since generative systems are often trained on generic datasets (Huang et al., 2023), their outputs risk being clichéd or homogenized, thereby diminishing artistic specificity. Rendering of the narration and speech bubble content for voiceover (also informed by MLLM analysis) would also play a key role in determining tone and context.

Despite potential pitfalls, the possibilities of a finely tuned system that auto-generates sound for a print comic, particularly one that operates within a binaural or spatial audio system, creates new possibilities for immersive texts. In particular, the rendering of the dialog, appropriate non-diegetic music and ambience in an immersive spatial audio environment, could provide opportunities to enhance a non-visual comic experience for the BLV community.

Addendum: Sites for further exploration—comics and voiceprint-enabled avatars.

Park speaks (2023) of the Naver Webtoon *Encountered* and how it immerses the reader by placing them inside the narrative space. The comic uses text input to capture the readers name which is then used in the dialog, and a camera photo that is used to create an avatar of the reader, turning them into a character that interacts in the story. The heightening of the immersive properties of personalised avatars in storytelling is well documented and can also humanize storytelling and sustain viewer attention. (Bell et al., 2018; Sakuma et al., 2023; Sylaiou et al, 2020).

It is not a great leap to imagine Webtoon projects that make use of the reader's voiceprint in this scenario. An emerging technology called voice cloning can be used to capture the

tonal nuances, pitching and phrasing of the reader by using captured voice recordings to train AI models such as RVC (Retrieval-based Voice Conversion) (Ghadekar et al., 2024; Szabó et al., 2025). This would allow the readers not only

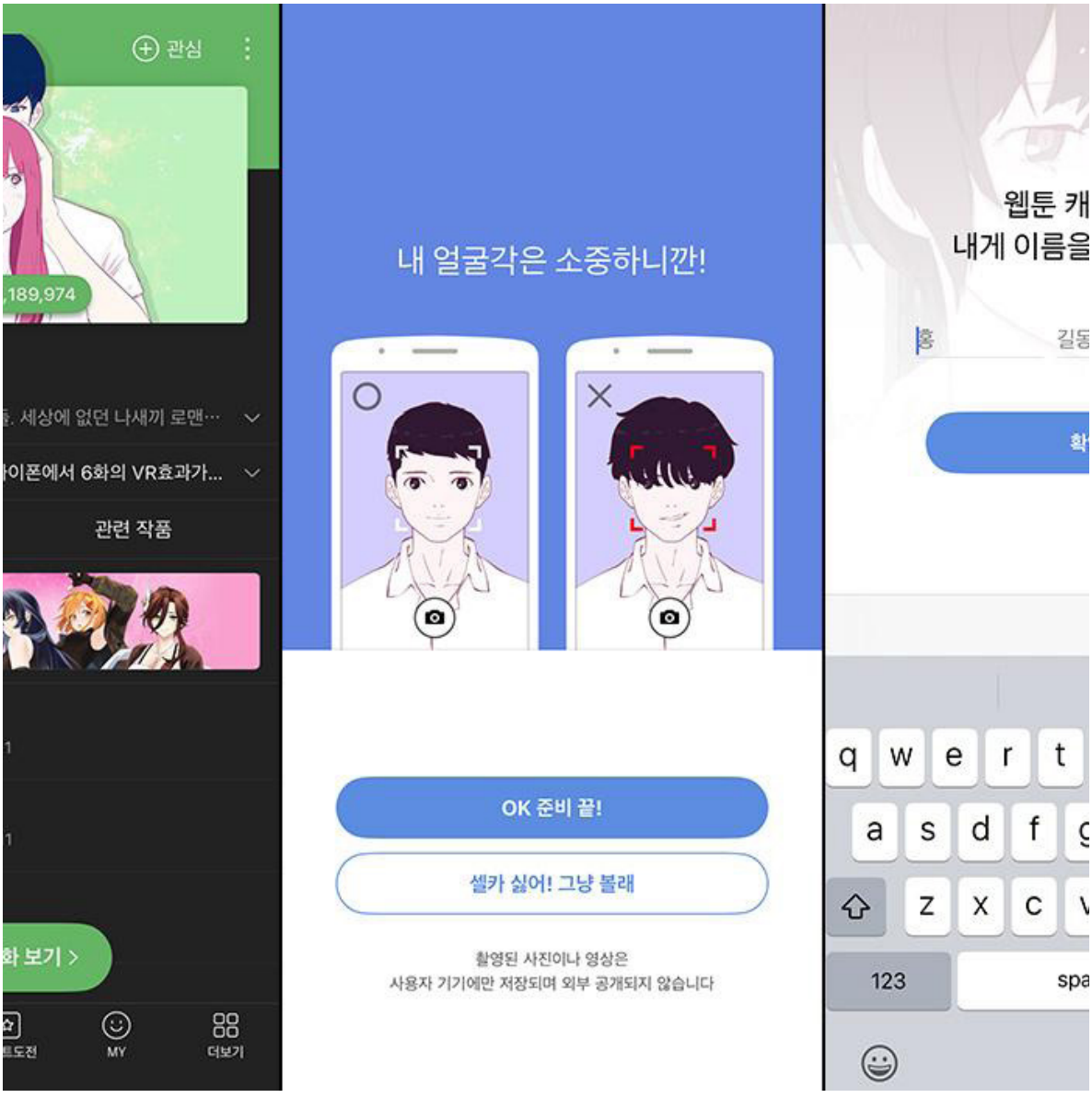

FIGURE 11 *Encountered* by Ha-Il Kwon personalizes the Webtoon experience via user text and photo input. Naver Webtoon Ltd, 2017

to see themselves in the story, but to hear themselves. This technology is already available in online platforms such as ACE Studio AI, MyVOcal.AI and ElevenLabs. The efficacy of this approach for enhancing immersion in digital comics is supported by arguments that the use of voice cloning in personalised storytelling promotes emotional closeness and engagement (Britto et el., 2025; Sathyan et al., 2025).

References

Adler, S. (2011). Silence in the graphic novel. *Journal of Pragmatics*, *43*(9), 2278–2285. https://doi.org/10.1016/j.pragma.2010.11.012

Aggleton, J. (2018). Defining digital comics: A British library perspective. *Journal of Graphic Novels and Comics*, *10*(4), 393–409. https://doi.org/10.1080/21504857.2018.1503189

Allen, T. (2014). *Economics of digital comics*. Indignant Media.

Altman, R. (1992). *Sound theory, sound practice*. Routledge.

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6077–6086. https://doi.org/10.48550/arXiv.1707.07998

Aja, D., & Fraction, M. (2023). *Hawkeye by Fraction & Aja: Omnibus*. Marvel Universe.

Asimakoulas, D. (2024). Your translated memory or mine? Remembering graphic novels in performed audio descriptions for The Cartoon Museum, London. *Journal of Graphic Novels and Comics, 15*(4), 491–507. https://doi.org/10.1080/21504857.2023.2277382

Barker, M. (1990). *Comics: Ideology, power and the critics*. St. Martin's Press.

Baroni, R. (2020). (Un)natural temporalities in graphic narratives. In Alber, J. & Richardson, B. (Eds.), *Unnatural narratology* (pp. 165-188). Ohio State University Press.

Bruhn, J. (2016). *The intermediality of narrative literature: Medialities Matter*. Palgrave Macmillan.

Bell, G., Ensslin, A., van der Bom, I., & Smith, J. (2018). Immersion in digital fiction. *International Journal of Literary Linguistics and Linguistic Theory, 7*(1), 1-22. https://doi.org/10.15462/ijll.v7i1.105

Benatti, F. (2024). *Innovations in Digital Comics: A popular revolution*. Cambridge University Press.

Bolter, J., & Grusin, R. (1999). *Remediation: Understanding new media*. MIT Press.

Britto, K.A., Harini, J., Shivanii, M., Karthik, P., Kalpana, B.N., Karthi, M. (2025). Audioverse: personalized storytelling using voice cloning. In Sharma, A., Rani, R. (Eds.), *Artificial intelligence and speech technology* (pp. 152-164). Springer. https://doi.org/10.1007/978-3-031-91331-0_11

Brooker, W. (2023). *Never ending Watchmen: Adaptations, sequels, prequels and remixes*. Bloomsbury Academic.

Cecchini, M. (2012, October 12). *Interview with Ben Wolstenholme, Madefire's co-founder*. Den of Geek. https://www.denofgeek.com/comics/interview-with-ben-wolstenholme-madefires-co-founder/

Chion, M. (1994). *Audio-vision: Sound on screen* (C. Gorbman, Trans.). Columbia University Press.

Chisholm, D. (2020). *Chasing the Bird: Charlie Parker in California*. Z2 Comics.

Coheed and Cambria. (2002). *The Second Stage Turbine Blade* [Album]. Defiance Records.

Cohn, N. (2005). Un-defining "Comics": Separating the cultural from the structural in comics. *International Journal of Comic Art. 7* (2). 236-248.

Cohn, N. (2010). The limits of time and Transitions: Challenges to theories of sequential image comprehension. *Studies in Comics*, *1*(1), 127–147. https://doi.org/10.1386/stic.1.1.127/1

Cohn, N. (2021). *Who understands comics?: Questioning the universality of visual language comprehension*. Bloomsbury Academic.

Coleman, O. (1983). Harmolodic democracy: Interview. *DownBeat*, *50*(10), 20–23.

Constine, J. (2012, March 11). *Marvel and Aurasma show off New Line of Augmented Reality Comics*. TechCrunch. https://techcrunch.com/2012/03/11/marvel-augmented-reality/

Cook, N. (1998). *Analysing musical multimedia*. Oxford University Press.

Daniels, E. (2016). *The Digital Comics Manifesto*. Screendiver. https://screendiver.com/digital-comics-manifesto/

Davis, B. (2023). Comics and multimodal storytelling. In M. Ahmed (Ed.), *The Cambridge companion to comics* (pp. 125–142). Cambridge University Press.

Deleuze, G. (1989). *Cinema 2: The time-image* (H. Tomlinson & R. Galeta, Trans.). University of Minnesota Press.

Dittmer, J. (2010). Comic book visualities: A methodological manifesto on geography, montage and narration. *Transactions of the Institute of British Geographers*, *35*(2), 222–236. https://doi.org/10.1111/j.1475-5661.2009.00376.x

Dittmar, J. (2019, August 4). Blind readers and comics: Reflecting on comics' storytelling from a different perspective. Comics Forum. https://comicsforum.org/2019/08/04/blind-readers-and-comics/

Dolby Laboratories (2018). Dolby Atmos Renderer guide (version 3.0) [Manual]. Dolby Laboratories. https://professional.dolby.com/siteassets/content-creation/dolby-atmos/dolby_atmos_renderer_guide.pdf

Echert, S. & Sanchez, T. (2004). *The Armory Wars* [Comic Book Series]. Evil Ink Comics.

Eco, U. (1979). *The role of the reader: Explorations in the semiotics of texts*. Indiana University Press.

Eisner, W. (1990). *Comics and sequential art: Principles and practice*. Poorhouse Press.

Elleström, L. (2010). *Media borders, multimodality and intermediality*. Palgrave Macmillan.

Elleström, L. (2021). The modalities of media II: An expanded model for understanding intermedial relations. In Elleström, L. (Ed.), *Beyond media borders*, Vol. 1, (pp. 3–92). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-49679-1

Fink, C. (2018). *Charlie Fink's metaverse - An AR enabled guide to VR & AR*. Cool Blue Media.

Fresnault-Deruelle, P. (1976). *La bande dessinée: Essai d'analyse sémiotique*. Hachette.

Gebhart, T. (2019). Digital Comics and Critical Librarianship. In Grimm, S., & Piepmeier, O. (Eds.), *Comics and critical librarianship: Reframing the narrative in academic libraries* (pp.25-39). Library Juice Press.

Ghadekar, P., Joshi, O., Barhate, J., Kundu, B., Nandeshwar, A., & Gujar, M. (2024, December). Recuperative Reference Search Aided Approach Towards Voice Cloning. In *2024 6th International Conference on Computational Intelligence and Networks (CINE)* (pp. 1-6). IEEE.

Gilles, S. (2020). Screen grammar for mobile frame media: The audiovisual language of cinematic virtual reality, case studies and analysis. In Knight-Hill, A. (Ed.), *Sound and image: Aesthetics and practices* (pp. 206-218). Routledge.

Gorbman, C. (1987). *Unheard melodies: Narrative film music*. Indiana University Press.

Groensteen, T. (2007). *The system of comics* (B. Beaty & N.

Nguyen, Trans.). University Press of Mississippi. (Original work published 1999).

Groensteen, T. (2012). The current state of French comics theory. *Scandinavian Journal of Comic Art*, *1*(1), 112–122. https://sjoca.com/wp-content/uploads/2012/06/SJoCA-1-1-Forum-Groensteen.pdf

Groensteen, T. (2013). *Comics and narration*. University Press of Mississippi.

Hague, I. (2011). *Hearing Comics: Sound and Sequential Art.* In: Joint International Conference of Graphic Novels, Bandes Dessinées and Comics, 5-8 July 2011, Manchester Metropolitan University.

Hague, I. (2014). *Comics and the senses: A multisensory approach to comics and graphic novels*. Routledge.

Hatfield, C. (2005). *Alternative comics: An emerging literature*. University Press of Mississippi.

Heer, J., & Worcester, K. (Eds.). (2009). *A comics studies reader*. University Press of Mississippi.

Helms, J. (2016). Potential Panels: Toward a theory of augmented comics. In Morey A., & Tinnel, J. (Eds.), *Augmented reality: Innovative perspectives across art, industry, and academia* (pp. 47-62). Parlor Press.

Holley, E. Jr. (2020, September 2). Bird in Hollywood: PW talks with Dave Chisholm. PublishersWeekly.com. https://www.publishersweekly.com/pw/by-topic/industry-news/comics/article/84240-bird-in-hollywood-pw-talks-with-dave-chisholm.html

Horton, I. & Gray, M. (2022). Modernism and comics revisited: Form and fragmentation in Alan Moore's "I Can Hear the Grass Grow". In *Art history for comics.* Palgrave Macmillan.

Huang, Q., Copet, J., Défossez, A., Synnaeve, G., & Adi, Y. (2023). *Make-an-audio: Text-to-audio generation with prompt-enhanced diffusion models*. arXiv preprint. https://doi.org/10.48550/arXiv.2301.12661

Ihde, D. (2008). *Listening and voice: Phenomenologies of sound*. State University of New York Press.

IGN Staff. (2008, July 19). *Warners presents motion comics*. IGN. https://www.ign.com/articles/2008/07/18/warners-presents-motion-comics

Iyyer, M., Manjunatha, V., Boyd-Graber, J., & Daumé III, H. (2017). *COMICS: a large-scale dataset of comics*. *h*ttps://doi.org/10.48550/arXiv.2306.17469

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.

Kashtan, A. (2018). Between pen and pixel: Comics, materiality, and the book of the future. Ohio State University Press.

Kleefeld, S. (2020). Webcomics. Bloomsbury Academic.

Koffka, K. (1935). Principles of Gestalt psychology. Harcourt. Internet Archive. https://archive.org/details/in.ernet.dli.2015.221555/page/n639/mode/2up

Krause, J., Johnson, J., Jia, Y., et al. (2017). A hierarchical approach for generating descriptive image paragraphs. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 880–888. https://doi.org/10.48550/arXiv.1611.06607

Kukkonen, K. (2013). *Contemporary comics storytelling*. University of Nebraska Press.

Kunzle, D. (1973). *The early comic strip: Narrative strips and picture stories in the European broadsheet from c.1450 to 1825*. University of California Press.

Langiulli, N., Calbi, M., Sbravatti, V., Umiltà, M. A., & Gallese, V. (2023). The effect of surround sound on embodiment and sense of presence in cinematic experience: A behavioral and HD-EEG study. *Frontiers in Neuroscience*, *17*, Article 1222472. https://doi.org/10.3389/fnins.2023.1222472

Lynk_XUS & Android k05 (2023). *Soulwinder* [Webtoon]. Webtoon. https://www.webtoons.com/en/thriller/soulwinder/list?title_no=2880.

Lastra, J. (2000). *Sound technology and the American cinema: Perception, representation, modernity*. Columbia University Press.

Lopez, M., Kearney, G., & Hofstadter, K. (2021). Enhancing audio description: Inclusive cinematic experiences through sound design. *Journal of Audiovisual Translation*, *4*(1). https://doi.org/10.47476/jat.v4i1.2021.154

MacDonald, J. (2023, September 29). *Korean dramas play a role in the booming popularity of Webtoons*. Forbes. https://www.forbes.com/sites/joanmacdonald/2023/09/14/korean-dramas-play-a-role-in-the-booming-popularity-of-webtoons/

McBride, S. (2008, July 18). *Web draws on comics' online shorts boost Batman*. Wall Street Journal. https://www.wsj.com/articles/SB121634908179464605

McDonald, H. (2021, April 29). *Madefire shuts down, leaving several digital comics apps in limbo*. The Beat. https://www.comicsbeat.com/madefire-shuts-down-leaving-several-digital-comics-apps-in-limbo/

McCloud, S. (1993). *Understanding comics: The invisible art*. Harper Perennial.

McCloud, S. (2000). *Reinventing comics: How imagination and technology are revolutionizing an art form*. Harper-Collins.

Mee, S. J., & Robinson, L. (Eds.). (2024). *Sound affects: A user's guide*. Bloomsbury.

Meskin, A. (2025). The philosophy of comics: What they are, how they work, and why they matter. *Journal of Aesthetics and Art Criticism*, 83(2), 187–194. https://doi.org/10.1093/jaac/kpaf018

Mikkonen, K. (2017). *The narratology of comic art*. Routledge.

Miodrag, H. (2013). *Comics and language: Reimagining critical discourse on the form*. University Press of Mississippi.

Morton, D. (2015). The unfortunates: towards a history and definition of the motion comic. *Journal of Graphic Novels and Comics*, *6*(4), 347–366. https://doi.org/10.1080/21504857.2015.1039142

Morton, D. (2017). Watched any good books lately?: The formal failure of the *Watchmen* motion comic." *Cinema Journal, 56*(2), 1–25.

Mougin, M. (2023, December 12). *Comic strips must find their place in teaching: Interview with Benoît Peeters*. Collège de France. https://www.college-de-france.fr/en/news/comic-strips-must-find-their-place-in-teaching

Ong, W. J. (1982). *Orality and literacy*. Taylor & Francis.

Overell, R. (2014). Intermediality and interventions: Applying intermediality frameworks to reality television and microblogs. *Refactory: A Journal of Entertainment Media*, *24*. https://refractory-journal.com/overell/

Panciroli, C., Fabbri, M., Luigini, A., Macauda, A., Corazza, L., Russo, V. (2023). *Augmented reality in Arts education*. In Nee, A. & Ong, S. (Eds.), Springer handbook of Augmented Reality (305-334). Springer. https://doi.org/10.1007/978-3-030-67822-7

Park, H. (2023). *Understanding Hallyu: The Korean wave through literature, Webtoon, and Mukbang*. Routledge.

Parker Records. (n.d.). *@parkerecords* [YouTube channel]. YouTube. Retrieved April 19, 2025, from https://www.youtube.com/@parkerecords

Peeters, B. (1991). *Case, planche, récit: lire la bande dessinée*. Casterman.

Pellitteri, M. (2019). The aural dimension in comic art. In M. Grimshaw-Aagaard, M. Walther-Hansen, & M. Knakkergaard (Eds.), *The Oxford handbook of sound and imagination*

(Vol. 1, pp. 510–548). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190460167.013.26

Peterson, R. (2009). The acoustics of Manga. In Heer, J. & Worcester, K. (Eds.), *A comics studios reader* (pp. 163-172). University Press of Mississippi.

Postema, B. (2013). *Narrative structure in comics: Making sense of fragments*. RIT Press.

Rajewsky, I. O. (2005). Intermediality, intertextuality, and remediation: A literary perspective on intermediality. *Intermédialités, (6)*, 43–64. https://doi.org/10.7202/1005505ar

Ram V. (2021). *Radio Apocalypse* (*1*) [Comic Book]. Vault Comics.

Ramprasad, R. (2023). *Comics for everyone: Generating accessible text descriptions for comic strips*. Microsoft. https://doi.org/10.48550/arXiv.2310.00698

Rancière, J. (2009). *The politics of aesthetics* (G. Rockhill, Trans.). Continuum.

Rigaud, C., Burie, J.-C., & Petit, S. (2024, July 11). *Toward accessible comics for blind and low vision readers* arXiv. https://arxiv.org/abs/2407.08248

Ritsos, P., Ritsos, D., & Gougoulis, A. (2011). Standards for Augmented Reality: a User Experience perspective. In C. Perey, J. de Smit & T. Adamek (Eds.), 2nd International Workshop on AR Standards (page numbers). International AR Standards Workshop. https://pdritsos.com/files/Ritsos-et-al-4th-AR-Standards.pdf

Sabharwal, A. (2015). *Digital curation in the digital humanities: Preserving and promoting archival and special collections*. Elsevier Science & Technology.

Sathyan, S., Midhun, M., Vishnu, V. S., Akash, R., & Rahul, R. (2025). An emotional support system using AI: Integration of voice cloning and multimodal capabilities. *International Journal of Innovative Research in Technology*, *11*(11), 6652–6656. https://ijirt.org/publishedpaper/IJIRT176783_PAPER.pdf

Sakuma, H., Takahashi, H., Ogawa, K., & Ishiguro, H. (2023). Immersive role-playing with avatars leads to adoption of others' personalities. *Frontiers in Virtual Reality*, *4*. https://doi.org/10.3389/frvir.2023.1025526

Satva. (2010). *Retinex* [AR Comic]. Metaverse One.

Schwartz, T. (1983). *Media, the second God*. Anchor Books.

Soykan, G., Yuret, D., & Sezgin, T. M. (2022, December 27). *A comprehensive gold standard and benchmark for comics text detection and recognition*. arXiv. https://doi.org/10.48550/arXiv.2212.14674

Sterne, J. (2012). *The sound studies reader*. Routledge.

Summers, T. (2015). "Sparks of Meaning": Comics, Music and Alan Moore. *Journal of the Royal Musical Association*, *140*(1), 121–162. http://www.jstor.org/stable/43741620

Sylaiou, S., Kasapakis, V., Gavalas, D., & (2020). Avatars as storytellers: affective narratives in virtual museums. *Personal and Ubiquitous Computing, 24*(5), 829–841.

Szabó, T., Žitný, R., Pšenáková, I., & Pšenák, P. (2025). Using Retrieval-based voice conversion in educational video materials. *R&E-SOURCE*, 352-362.

The Lakes International comic art festival. (2020). *Benoît Peeters - Time, space and narration, from Winsor McCay to Richard McGuire* [Video]. YouTube. https://www.youtube.com/watch?v=6j9lOgaNSvs

Timothy Truman & the Dixie Pistols. (1987). *Marauders* [Album]. Eclipse Records.

Trabado, J. M. (2022). Reseña de "Sistema de la historieta" de Thierry Groensteen. *CuCo, Cuadernos de Cómic*, (18). https://doi.org/10.37536/cuco.2022.18.1955

Truman, T. (1987). *Scout (19)* [Comic Book]. Eclipse Comics.

Tucker, H. J., Minichiello, M., & Roxburgh, M. (2018). The motion comic: Neither something nor nothing." *Journal of Communication and Media Studies, 3*(1), 13–24.

Wertheimer, M. (1923). *Untersuchungen zur Lehre von der Gestalt II* [Laws of organization in perceptual forms]. *Psychologische Forschung, 4*, 301–350.

Wincott, A., Martin, J., & Richards, I. (2021). Telling stories in soundspace: Placement, embodiment and authority in immersive audio journalism. *Radio Journal: International Studies in Broadcast & Audio Media*, *19*(2), 123–141.

Wolf, W. (2015). Literature and music: Theory. In Rippl, G. (Ed.), *Handbook of intermediality* (pp. 459–474). De Gruyter.

Young, S. (2023). *I hate fairyland* (5) [Comic Book]. Image Comics.