

# The Artistic Editing of Ambient Sound and the Construction of Emotional Tension in Urban Vlogs

Xin Zhou<sup>1</sup> & Pengsong Gao<sup>1</sup>

<sup>1</sup> Nanchang University, Jiangxi, China

Correspondence: Xin Zhou, Nanchang University, Jiangxi, China.

doi:10.63593/AS.2709-9830.2025.05.006

## Abstract

This paper explores the aesthetic, emotional, and narrative roles of ambient sound in urban vlogs, situating this practice within broader East Asian media cultures and global soundscape theory. Through a multi-faceted analysis that integrates case studies, cross-cultural comparison, and theoretical synthesis, the study reveals how vloggers in China transform everyday soundscapes—such as the clatter of street vendors, the murmur of public spaces, and the textures of weather—into immersive emotional experiences. Ambient sound is not merely a background element in these vlogs; it functions as a powerful narrative device that shapes viewer engagement, evokes nostalgia, and constructs sensory spaces that bridge individual memory and collective identity. Drawing upon frameworks from soundscape theory, emotional geography, and multimodal narrative, the paper demonstrates how ambient sound editing in vlogs reflects a unique form of sonic poetics: a blending of cultural memory, personal storytelling, and aesthetic curation that extends beyond visual storytelling alone. The analysis highlights the participatory dynamic of audience reception, showing how viewers interpret, respond to, and emotionally connect with these curated soundscapes. Ultimately, the study positions ambient sound in vlogs as a key site for understanding contemporary digital storytelling, emotional engagement, and the evolving relationship between sound, space, and culture.

**Keywords:** ambient sound, vlogs, soundscape theory, emotional geography, multimodal narrative, audience reception

## 1. Introduction

Urban vlogs have rapidly emerged as an evocative medium at the intersection of digital culture, personal expression, and collective urban memory. These vlogs, widely distributed on platforms such as Bilibili, Douyin, and Weibo, are not merely visual diaries; they are complex multimodal texts that reflect the rhythms, contradictions, and atmospheres of rapidly evolving metropolises such as Shanghai, Guangzhou, Chengdu, and Wuhan. Within this emerging genre, ambient sound has evolved into a critical aesthetic and affective tool, used not only to enrich spatial realism but to create emotionally charged narratives that resonate deeply with viewers navigating similar urban experiences. Sound in vlogs—particularly ambient audio like street chatter, subway rumble, bicycle bells, or sudden silence—functions on both a literal and symbolic level. The editing of this soundscape shapes mood, guides audience perception, and articulates a narrative tension that is often more affective than plot-driven. Unlike traditional film scoring, which relies on composed music to evoke emotion, vloggers in urban China often embrace diegetic sound: real-world, environmental audio captured or constructed to mirror psychological states or urban anxieties. This conscious manipulation transforms ambient noise into an artistic medium of emotional storytelling.

As China's urbanization accelerates and social fabric becomes increasingly mediated by digital platforms, vlogs serve as both an archive of everyday experience and a performative expression of urban subjectivity. Drawing from theories of digital media aesthetics, soundscape studies, and emotional geography, this essay examines how

ambient sound editing in vlogs contributes to the construction of emotional tension. This includes the ways vloggers modulate tempo, isolate noise, layer resonance, or oscillate between cacophony and silence to dramatize urban isolation, nostalgia, excitement, or alienation. The study also situates this practice within broader East Asian ambient media traditions, highlighting how sonic atmosphere becomes a tool for both identity construction and emotional catharsis. This inquiry not only deepens our understanding of vlog-based narrative aesthetics but also explores how ambient sound editing functions as a creative response to the lived complexity of contemporary cities. Through a close examination of vlog samples, sound theory, and cultural analysis, this essay argues that ambient sound is not mere background—but a dynamic, edited foreground that animates and intensifies the emotional lifeworld of the digital urban flâneur.

## 2. Ambient Sound as a Narrative Device

In the realm of urban vlogs, ambient sound transcends its traditional role as mere background noise, emerging as a pivotal narrative element that shapes viewers' emotional and cognitive engagement. Vloggers meticulously capture and manipulate environmental sounds—ranging from the rhythmic clatter of Mahjong tiles in a Chengdu teahouse to the distant hum of traffic in Shanghai—to construct immersive soundscapes that reflect the complexities of urban life in contemporary China. This section delves into the multifaceted role of ambient sound in vlogs, exploring how it functions as a narrative device that conveys emotion, constructs space, and engages with socio-cultural discourses.

### 2.1 *The Role of Environmental Audio: From Documentation to Emotional Poetics*

Ambient sound in vlogs serves not only to document the sonic environment but also to evoke emotional responses and convey nuanced narratives. Drawing upon R. Murray Schafer's concept of the "soundscape," vloggers curate auditory experiences that encapsulate the cultural, social, and historical specificities of place. For instance, the juxtaposition of traditional instruments like the erhu or guzheng with the background hum of high-speed trains symbolizes the tension between heritage and modernity, prompting introspection and evoking a sensory memoryscape that transcends visual storytelling alone.

These ambient sounds function as sonic metaphors for broader social phenomena: the anxiety of urbanization is encoded in the ceaseless drone of traffic; the fragility of interpersonal connections is mirrored in the fleeting strains of an old pop ballad heard from a passing bicycle radio; the endurance of cultural identity is expressed in the persistent, though faint, echoes of folk melodies in crowded public squares. The vlogger, then, becomes an emotional cartographer, mapping not only the geography of the city but also the emotional undercurrents that flow beneath its surface.

### 2.2 *The Rhythmic Editing of Sound: Sonic Pacing, Tension, and Emotional Flow*

Editing ambient sound is an art of temporal and emotional modulation. Vloggers engage in sonic sculpting—intensifying, prolonging, or silencing specific sounds to guide audience attention, evoke affect, and structure narrative flow. This is not unlike film sound design, yet vloggers adapt these techniques to the unique affordances of vlogging: episodic formats, first-person perspectives, and an intimate proximity to everyday life.

For example, in capturing the chaotic vibrancy of a morning market, a vlogger might amplify the sharp crack of a cleaver slicing through pork, the rhythmic bargaining of a fishmonger, and the clatter of plastic bags—all mixed into a dense sonic texture that conveys urgency and sensory overload. In contrast, they may transition abruptly to a shot of an empty stairwell, where the only audible sound is the faint hum of an air conditioner. This sonic contrast—from density to sparsity—serves as a rhythmic breath, a moment of pause that allows viewers to reflect, recalibrate, and emotionally reset. Such modulation of acoustic density mirrors what Michel Chion describes as "empathetic sound"—sound that directly aligns with the emotional tone of the visual narrative.

Silence, too, plays a crucial narrative role. In the soundscape of vlogs, sudden silences—such as the moment after a subway train departs or the lull between raindrops during a downpour—function as negative space, emphasizing absence, isolation, or unresolved tension. These sonic voids create emotional punctuation marks, signaling shifts in mood or perspective and encouraging viewers to linger in a moment of introspection. Vloggers frequently employ micro-synchronization of sound with image—matching the rhythm of footsteps to scene cuts, or aligning the revving of a scooter with a camera pan—to achieve affective resonance. This rhythmic alignment transforms vlogs into multimodal essays, blending the documentary gaze with poetic subjectivity. The audience is drawn into a felt experience, where sound guides not only what is seen but how it is emotionally processed.

### 2.3 *Ambient Sound as Subtle Activism: Claiming the Right to Be Heard*

Beyond aesthetics, the ambient soundscapes curated by vloggers often carry implicit socio-political critique. In an urban landscape increasingly shaped by surveillance, standardization, and commercial homogenization, the decision to foreground local dialects, vernacular music, and unfiltered street noise is an act of acoustic resistance.

By amplifying the messy, unpredictable, and diverse sounds of the city—sounds that are often sanitized or erased in state media representations—vloggers reclaim the right to be heard in a rapidly transforming urban environment. In this sense, ambient sound becomes a site of memory, identity, and subtle defiance. The small, everyday noises—children playing in alleyways, the hawking of street vendors, the whistle of a kettle in a communal kitchen—become archives of the ordinary, countering the grand narratives of progress and control. These sounds, carefully curated and edited, offer a counterpoint to the spectacle of modernity, reminding viewers of the human, intimate, and affective textures that persist beneath the city's surface.

#### *2.4 Exemplifying the Narrative Power of Ambient Sound*

To illustrate the theoretical concepts discussed, we can examine specific case studies of urban vlogs that effectively utilize ambient sound as a narrative device. *Midnight Noodles in Chengdu*: In one vlog, the creator documents a late-night visit to a noodle stall in Chengdu. The ambient sounds—the sizzling of oil, the clatter of bowls, the murmur of patrons—are prominently featured, creating an immersive atmosphere that transports viewers to the scene. The vlogger employs minimal narration, allowing the sounds to convey the warmth and vitality of the nocturnal food culture. The rhythmic editing of these sounds, synchronized with the visual pacing, enhances the sensory experience and evokes a sense of comfort and nostalgia. *Morning Commute in Beijing*: The vlog captures the experience of commuting during rush hour in Beijing. The ambient sounds—train announcements, footsteps, conversations—are layered to reflect the bustling energy and underlying tension of the urban environment. The vlogger strategically uses moments of silence, such as the pause between train arrivals, to highlight the contrast between movement and stillness, crowd and solitude. This sonic narrative mirrors the emotional fluctuations of the commuter experience, from anticipation to fatigue. *Rainy Afternoon in Suzhou Gardens*: In a vlog, the creator explores a traditional garden in Suzhou during a rainstorm. The ambient sounds—raindrops on leaves, distant thunder, the rustle of bamboo—are meticulously recorded and edited to emphasize the tranquility and introspection associated with the setting. The vlogger's use of binaural recording techniques enhances the spatial realism of the soundscape, inviting viewers into a meditative state. This approach aligns with the aesthetic principles of garden design, where sound plays a crucial role in shaping the sensory experience.

#### *2.5 Soundscapes and Chinese Aesthetics*

The use of ambient sound in vlogs is deeply rooted in cultural aesthetics and philosophies. Traditional Chinese art forms, such as landscape painting and poetry, emphasize the harmony between humans and nature, often evoking sensory experiences through imagery and rhythm. This sensibility extends to the auditory realm, where sounds are not merely background elements but integral components of the aesthetic experience. The concept of “jing”, meaning “realm” or “state,” is central to Chinese artistic expression. In vlogging, ambient sound contributes to the creation of a “jing” by immersing viewers in a specific sensory and emotional environment. For instance, the sound of rain in a garden scene evokes a contemplative mood, aligning with the Daoist appreciation for natural cycles and tranquility. The integration of ambient sound reflects the Confucian value of “li”, or ritual propriety, by acknowledging and respecting the sonic characteristics of a place. By carefully capturing and presenting these sounds, vloggers demonstrate reverence for the cultural and historical significance embedded in the urban soundscape.

Ambient sound in Chinese urban vlogs functions as a powerful narrative device that enriches storytelling by conveying emotion, constructing space, and engaging with cultural discourses. Through meticulous recording and editing, vloggers transform everyday sounds into immersive soundscapes that resonate with viewers on multiple levels. These practices not only enhance the aesthetic quality of vlogs but also contribute to the preservation and appreciation of China's diverse urban environments. As digital media continues to evolve, the role of ambient sound in narrative construction will likely expand, offering new opportunities for creative expression and cultural engagement. Future research could explore the impact of emerging technologies, such as virtual reality and spatial audio, on the production and consumption of ambient soundscapes in digital storytelling.

### **3. Emotional Tension and Urban Experience**

In the rapidly urbanizing landscapes of contemporary China, the sensory experiences of city dwellers are in constant flux. Among these sensory modalities, sound plays a pivotal role in shaping the emotional and psychological landscapes of urban life. Vloggers, as chroniclers of everyday experiences, have increasingly harnessed ambient sound as a narrative device to convey the complexities of urban existence. This section explores how ambient sound in vlogs constructs emotional tension and reflects the multifaceted urban experience, drawing upon theoretical frameworks from sound studies, urban sociology, and cultural geography.

#### *3.1 The Soundscape of Urban China*

The concept of the “soundscape,” as introduced by R. Murray Schafer, refers to the acoustic environment as

perceived by humans. Schafer distinguishes between “hi-fi” soundscapes, characterized by clarity and low ambient noise, and “lo-fi” soundscapes, marked by high levels of ambient noise and reduced acoustic clarity. urban environments often exemplify the latter, with dense populations and incessant construction contributing to a cacophonous auditory backdrop.

However, vloggers selectively curate these soundscapes, emphasizing certain sounds while muting others to craft a particular emotional narrative. This practice aligns with the notion of “affective atmospheres,” which refers to the shared, spatially distributed feelings that permeate social life. By manipulating ambient sound, vloggers create affective atmospheres that resonate with viewers, evoking emotions ranging from nostalgia to anxiety.

### 3.2 *Sonic Cartographies of the Chinese City*

The emotional tension constructed through ambient sound in urban vlogs cannot be dissociated from the complex spatial contradictions of China’s post-socialist urbanization. Vloggers navigate a fragmented urban fabric—where the glitter of glass-clad skyscrapers towers over crumbling hutongs, where high-speed trains rush past rusting factory districts, and where mega-malls coexist with ancestral temples. These physical juxtapositions create acoustic dissonances: the heavy bass of a nightclub bleeding into the quiet chime of a Buddhist bell, the metallic clank of construction equipment merging with the crackle of fireworks during Spring Festival. Vloggers capture these sounds not as passive observers, but as curators of urban atmospheres, selectively foregrounding or muting them in post-production to construct emotional landscapes.

The editing of ambient sound mediates the relationship between external space and internal feeling. For example, in a scene where a vlogger walks alone down a rain-slicked alley in Beijing, the rhythmic amplification of footsteps—each echo stretching into a soft reverb—creates a sense of isolation, magnifying the emptiness of the surrounding environment. The distant wail of a siren, layered subtly under the ambient track, introduces an element of unease, evoking both the literal presence of authority and the psychological undercurrent of urban surveillance.

The persistent hum of a public address system in a Shenzhen metro station may seem mundane, but when emphasized through selective editing—its flat, impersonal cadence rising just above the surrounding chatter—it becomes a symbol of systemic control and the erasure of individuality. This practice of sonic framing allows vloggers to transform ordinary sounds into emotionally charged cues, turning everyday spaces into affective micro-dramas that reflect broader tensions: between freedom and regulation, nostalgia and progress, intimacy and alienation.

These choices contribute to what cultural geographers call “affective atmospheres”—the shared, spatially distributed feelings that permeate social life. The compressed rush of motorbikes, the faint strains of a love song leaking from a convenience store, the sudden hush after a burst of fireworks—all these sounds become emotional textures woven into the fabric of urban experience. Through careful sound design, vloggers invite viewers to *feel* the city: the claustrophobia of crowded subway cars, the melancholy of twilight streets, the thrill of bustling night markets. Each auditory decision, from the softening of background noise to the insertion of an isolated soundbite, is a brushstroke in the emotional portrait of the city.

### 3.3 *Sonic Remembrance in Post-Traumatic Urbanism*

In cities marked by collective trauma—such as Wuhan in the aftermath of the COVID-19 lockdown—ambient sound assumes an even greater psychological and symbolic significance. Here, soundscapes are not only tools of narrative tension but also mediums of healing, resilience, and cultural continuity. W. Zhang documents how vloggers in Wuhan reintroduced familiar sounds into their videos as a way of reclaiming public space and rebuilding collective memory. The click of bicycle bells, once muted during the eerie quiet of lockdown, now rings louder, symbolizing a return to normalcy. The sizzle of street food stalls, the rhythmic cadence of tai chi music in parks, and the chatter of vendors hawking fruit become not just sound effects but auditory testaments to survival and resilience.

This sonic nostalgia is carefully crafted through editing. Vloggers often enhance certain sounds—raising their volume slightly, adding subtle reverb, or looping them briefly to create a lingering presence—thus transforming mundane noises into sonic monuments. These soundscapes become memory-objects, connecting viewers to a shared emotional past. When audiences hear the faint strains of a familiar pop song drifting from a café in Wuhan or the murmur of elderly neighbors gossiping outside a convenience store, they are not just recalling a specific scene; they are participating in a ritual of collective healing, re-weaving the broken threads of daily life into a tapestry of belonging.

This affective work resonates with anthropologist Kathleen Stewart’s concept of “ordinary affects”—the subtle, often overlooked feelings that circulate in everyday environments. Through ambient sound, vloggers amplify these affects, allowing moments of tenderness, loss, hope, or anxiety to surface within the flow of visual narrative. The result is an emotional landscape that is both personal and collective, where viewers can feel the

weight of urban alienation but also the warmth of shared resilience.

Ambient sound in vlogs often blurs the line between documentary and affective fiction. For instance, a vlogger may overlay the sound of a crackling vinyl record over footage of a modern shopping mall, creating a temporal dissonance that evokes longing for a slower, analog past in the face of hyper-consumerist modernity. Or they might add soft, echoing footsteps to a scene of an empty street at dawn, imbuing the silence with a sense of quiet anticipation or existential uncertainty. These aesthetic choices are not random; they are emotional scripts that guide viewers through the complex, often contradictory feelings that define contemporary urban life.

The editing of ambient sound in vlogs is not just a technical layer of production—it is a deeply affective practice that transforms raw sensory material into narratives of place, memory, and emotion. It is a form of sonic storytelling that captures the emotional aftershocks of rapid urban change and offers viewers a space for reflection, empathy, and collective identification.

#### 4. Comparative and Cross-Cultural Perspectives

The use of ambient sound as a narrative and affective tool in vlogs is part of a broader East Asian media ecology shaped by cultural aesthetics, historical traditions, and technological advancements. To fully understand the significance of ambient sound editing in vlogs, we must compare it with parallel practices in Japanese, Korean, and other global media cultures. This comparative lens illuminates how soundscapes are not just a backdrop but a rich emotional layer that conveys mood, memory, and atmosphere.

##### 4.1 East Asian Aesthetic Lineages

Japanese ambient media, especially its connection to concepts like wabi-sabi and ma, emphasizes natural soundscapes: rain on rooftops, waves lapping against the shore, birdsong in a quiet forest. These sounds, often presented in continuous loops, create a meditative, introspective atmosphere. The Japanese aesthetic leans toward creating tranquil soundscapes that evoke timelessness and serenity.

By contrast, vlogs embrace a sonic realism deeply embedded in the rhythms of urban life. While natural elements like the rustling of leaves or the trickle of a river may appear, they are often juxtaposed with the vibrancy of city soundscapes: the crackle of a street food wok, the rhythmic clatter of Mahjong tiles, the gentle murmur of dialects in a market, or the faint melodies of traditional instruments woven into the hum of modernity. These sounds do not merely document but evoke the tension between heritage and change, community and individuality.

Table 1.

Aspect	Japanese Ambient	Chinese Vlogs
Dominant Sounds	Rain, wind, nature, subtle reverb	Street vendors, traffic, traditional instruments
Emotional Register	Calm, introspective, nostalgic	Vibrant, bittersweet, layered
Aesthetic Principle	Wabi-sabi, Ma, nature	Sonic realism, urban immersion, cultural memory
Editing Technique	Long loops, minimal cuts	Rhythmic pacing, layering, contrasts

##### 4.2 Sound as Emotional Archive

Ambient sound in Chinese vlogs is often an emotional archive: it connects viewers to sensory memories of place, season, and shared experiences. A recording of bicycles rattling over cobblestones may evoke summer evenings in a small alley; the buzz of cicadas calls forth the heat of a July afternoon. Vloggers consciously select and emphasize these sounds—sometimes isolating them, sometimes layering them—to create a sonic tapestry that invites the audience into a shared affective space.

This differs from Japanese ambient streams, where the goal is often to cultivate a universal sense of calm. Chinese vlogs, in contrast, are more specific and textured: they do not aim to blur time and space but to situate the viewer in a precise moment, whether it is the bustle of a Lunar New Year market, the echo of footsteps on an old stone bridge, or the soft chatter of a family dinner.

Korean ASMR culture, known for its precise sound design (the crackle of paper, the fizz of a soda can), shares some aesthetic overlap but often centers on controlled, stylized sound events. Chinese vlogs, by contrast, embrace ambient imperfection: overlapping voices, the uneven rhythm of daily life, the unpredictable textures of weather and street sounds.

##### 4.3 Layered Storytelling: Ambient Sound within Multimedia Narratives

Chinese vlogs rarely present ambient sound in isolation. It is woven into a multimedia canvas: slow-motion shots

of lanterns swaying in the wind; close-ups of food sizzling on a grill; poetic captions in flowing script; the soft strains of a guzheng rising beneath the sound of rain. This layering creates an experience that is not purely visual or auditory but multisensory.

For example, a scene of an old woman tending her garden may feature the rustle of leaves, the gentle clinking of her tools, and a whispered narration reflecting on the passage of time. The viewer is immersed not just in an image but in a felt atmosphere, where sound evokes emotion, time, and place.

Table 2.

Modal Layer	Chinese Vlogs
Visual	Cityscapes, nature, food, daily rituals
Ambient Sound	Urban noise, traditional music, environmental textures
Text	Poetry, reflections, cultural idioms
Music	Folk instruments, electronic beats, nostalgic melodies

#### 4.4 Cross-Cultural Flow and Global Resonance

While rooted in local cultures, Chinese vlogs resonate with global audiences, in part due to the universality of sound as a medium of emotional expression. Viewers from different cultures may not understand a dialect spoken in a market, but they will recognize the emotional tenor of laughter, the gentle hum of a quiet evening, or the shared hush of a first snowfall. At the same time, Chinese vlogs contribute to a global conversation about sound and mood: they draw on cinematic techniques (sound bridges, fades, sudden silences), ASMR practices (close-up audio, binaural recording), and digital aesthetics (looped textures, sound overlays). The result is a hybrid soundscape: a blend of the hyperlocal and the transnational, where a rainstorm in Hangzhou or a festival in Chengdu becomes part of a shared digital sound archive.

Ambient sound in Chinese vlogs functions as a sonic poetics—a way of feeling and remembering, of situating oneself in the flow of daily life, of sensing the tension between the fleeting and the enduring. Through their sound editing practices, vloggers invite viewers into a world that is at once immediate and timeless, particular and universal. This practice reflects a broader East Asian aesthetic sensibility: an appreciation for the ephemeral, for small moments, for the beauty of imperfection. At the same time, it speaks to the global potential of sound as a medium: a way to bridge cultures, evoke shared emotions, and create immersive emotional landscapes in the digital age.

#### 5. Ambient Sound in Scholarly Frameworks

While the aesthetic and emotional power of ambient sound in vlogs has been explored through case studies and cross-cultural analysis, it is crucial to contextualize these practices within broader theoretical frameworks. This section synthesizes key concepts from sound studies, cultural geography, media anthropology, and narrative theory, positioning this study within a multi-disciplinary discourse.

At the foundation lies R. Murray Schafer's concept of the soundscape: the total acoustic environment as perceived by humans. Schafer distinguishes between hi-fi and lo-fi soundscapes—an urban street in Chengdu, with overlapping traffic noise, dialects, and vendor calls, exemplifies a lo-fi environment where sounds compete and blend. Vloggers, through artistic editing, engage in a form of acoustic ecology, selectively amplifying or muting elements to curate an emotional experience. This resonates with Schafer's idea of the "schizophonia"—the separation of sound from its source in modern media culture. Vlogs, by re-contextualizing ambient sounds, often dislocate them from their original time and space, transforming them into emotional cues and narrative devices.

Building upon Schafer, sound geography (Rodaway, 1994) suggests that sound is not merely a sensory experience but a spatial one. In vlogs, ambient sound becomes a tool of spatial narration: the echo of footsteps in a narrow alley conveys confinement; the distant murmur of a festival creates a sense of community; the isolated clang of a closing gate evokes finality. The vlogger, by manipulating these spatial sound cues, invites viewers to inhabit a sonic map of the urban environment, constructing emotional geographies that mirror the lived experience of Chinese cities.

Emotional geography (Anderson, 2009) emphasizes how emotions are spatially embedded. The nostalgia evoked by the crackle of an old radio in a vlog, or the sense of serenity elicited by the sound of flowing water in a garden, illustrates how ambient sound is not neutral—it is a carrier of collective memory and emotional resonance. Vlogs use sound to articulate feelings of belonging, loss, or curiosity, translating the textures of

everyday life into shared emotional experiences.

The concept of multimodal narrative (Kress & van Leeuwen, 2001) highlights how meaning in media emerges from the interplay of multiple semiotic modes. In Chinese vlogs, ambient sound interacts dynamically with visuals (e.g., neon-lit streets), text (e.g., poetic captions), and music (e.g., traditional melodies). This layering of modalities enriches the narrative, creating immersive affective environments where sound guides emotional interpretation, anchors spatial awareness, and deepens the sensory impact of visual storytelling.

Finally, the theory of affective atmospheres (Anderson, 2009; Bissell, 2010) underscores how sound contributes to the emotional texture of a space. Vloggers do not merely record places; they design atmospheres through sound: the bustle of a wet market conveys vitality, while the muffled hush of a library scene evokes introspection. By editing ambient sound, they sculpt felt spaces, shaping how viewers perceive and emotionally engage with the urban world.

Incorporating these theoretical perspectives allows for a richer interpretation of ambient sound editing in vlogs. It situates the practice not only as an artistic technique but as a form of emotional cartography, cultural storytelling, and sensory immersion. This synthesis sets the stage for the concluding reflections on how these soundscapes resonate within and beyond cultural boundaries.

## 6. Audience Reception and Interpretation

While much of this study has focused on the artistic strategies employed by vloggers in constructing ambient soundscapes, it is equally important to examine how audiences perceive, interpret, and emotionally engage with these sound environments. This section draws upon audience comments, viewer responses, and content analysis to understand the affective impact of ambient sound on viewers and how these sonic elements contribute to the broader appeal of vlogs.

### 6.1 Sensory Engagement

Audience comments on platforms like Bilibili and Xiaohongshu frequently highlight the immersive power of sound in vlogs. Viewers often describe a sense of “being there,” evoked by the precise layering of ambient sounds such as the crackling of oil in a street stall or the rustling of leaves in an alley. Comments like “I could almost smell the food through the sound!” or “This rain sound brings back memories of my childhood summers” illustrate how ambient sound stimulates multi-sensory engagement, blurring the line between viewing and experiencing. These responses suggest that ambient sound serves as a form of emotional transportation—a conduit that allows viewers to imaginatively inhabit the spaces depicted in vlogs. The gentle clink of chopsticks in a noodle shop, the soft patter of rain on tiled roofs, or the distant hum of a city bus become not just background details but emotional triggers, sparking nostalgia, curiosity, or even longing in viewers.

### 6.2 Emotional Labels in Audience Discourse

Content analysis of viewer comments reveals recurring emotional descriptors attached to specific soundscapes. Words such as “nostalgic,” “healing,” “soothing,” and “warm” appear frequently in response to vlogs that feature gentle, familiar sounds like the clink of teacups or the distant chatter of markets. Conversely, vlogs with dense, chaotic soundscapes—such as crowded urban streets—elicit emotions like “exciting,” “anxious,” or “alive.” This pattern suggests that ambient sound not only reflects the emotional tone of a vlog but also shapes the audience’s affective interpretation of place and narrative.

Table 3.

Sound Element	Emotional Response (from Audience Comments)
Rainfall on rooftops	Calm, nostalgic, comforting
Street vendors shouting	Lively, dynamic, community-centered
Metro announcements	Hectic, tense, overwhelming
Crackle of a fire	Cozy, warm, home-like
Footsteps in an alleyway	Lonely, reflective, melancholic

These associations indicate that viewers not only passively consume vlogs but actively interpret and internalize ambient sound as part of their personal affective landscapes.

### 6.3 The Participatory Aesthetic: Viewer as Co-Author

Audience responses also reflect a participatory aesthetic, where viewers co-construct meaning through comments, suggestions, and personal associations. For example, a viewer might write, “This sound reminds me

of my hometown,” while another might overlay their own recording or create a remix using the vlog’s ambient track. These interactions extend the affective resonance of the vlog beyond the creator’s intent, demonstrating how ambient sound fosters shared emotional worlds in digital spaces. Some creators even solicit feedback, asking viewers how a certain sound made them feel or inviting suggestions for future sound-focused vlogs. This dialogic process blurs the boundary between content creator and audience, transforming ambient sound from a passive narrative layer into a collaborative affective medium.

By incorporating audience reception into the analysis, we complete the interpretive loop of production—text—reception. Ambient sound in vlogs is not merely an artistic device; it becomes a collaborative emotional field, where creators and audiences together engage in shared listening, remembering, and feeling. This participatory dynamic highlights the importance of considering not just the creation but also the reception and circulation of ambient sound in the evolving aesthetics of digital media. Understanding how audiences respond to these sounds opens new avenues for research into cross-cultural affect, media literacy, and the evolving role of sound in global digital storytelling. Future studies might investigate how ambient sound influences viewer retention, emotional identification, and the broader cultural resonance of vlogs in a transnational media landscape.

## 7. Conclusion

This study has examined the narrative and affective role of ambient sound in urban vlogs, revealing its centrality in shaping emotional tension, constructing sensory space, and expressing cultural memory. Through a multi-layered analysis of aesthetic lineage, sound as emotional archive, multimodal storytelling, theoretical synthesis, and audience reception, we find that ambient sound operates not merely as background noise, but as an intentional and nuanced narrative tool that deepens viewer immersion and affective resonance.

Vloggers engage in a sophisticated practice of sonic curation. Drawing from a uniquely East Asian aesthetic sensibility while responding to the contemporary dynamics of urban life, they transform ordinary soundscapes into rich emotional textures. Unlike the minimalist loops common in Japanese ambient media or the polished clarity of Western ASMR, the ambient sounds in vlogs embrace realism, complexity, and hybridity—fusing dialects, environmental noise, and music into an emotionally compelling sound environment.

Ambient sound in vlogs intersects with frameworks of soundscape theory, emotional geography, and multimodal narrative. It enables what we have called “sonic storytelling”—a process through which vloggers construct place, evoke mood, and stimulate memory not through words alone, but through rhythm, tone, and atmosphere. By embedding viewers into the felt world of urban China, these vlogs produce not only a visual document of the city, but a deeply affective sonic archive. The audience, as demonstrated through reception analysis, is not a passive listener. Viewer comments often describe emotional triggers—nostalgia, calm, warmth, melancholy—attached to specific auditory cues. The ambient sound becomes a shared affective field through which creators and audiences collectively experience, recall, and reinterpret the emotional weight of everyday environments.

Ambient sound should be seen not as a supporting feature of vlog production, but as a central mechanism of storytelling. It is a modality of digital authorship that allows vloggers to express not just where they are, but how it feels to be there. Future research may further explore the algorithmic and technological dimensions of sound in digital media: how platforms shape sound visibility, how spatial audio (3D/VR) might extend sonic immersion, and how machine learning tools affect the aesthetics of vlog editing. But at the core, the study of ambient sound in vlogs reminds us of something profoundly human: that we listen not only to hear, but to remember, to feel, and to connect.

## References

- Anderson, B., (2009). Affective atmospheres. *Emotion, Space and Society*, 2(2), 77-81. <https://doi.org/10.1016/j.emospa.2009.08.005>
- Bissell, D., (2010). Passenger mobilities: Affective atmospheres and the sociality of public transport. *Environment and Planning D: Society and Space*, 28(2), 270-289. <https://doi.org/10.1068/d3909>
- Dancyger, K., (2018). *The technique of film and video editing: History, theory, and practice*. Taylor & Francis.
- Gao, W., Tian, Y., Huang, T., & Yang, Q., (2010). Vlogging: A survey of videoblogging technology on the web. *Proceedings of the 18th ACM International Conference on Multimedia*, 1-10. DOI:10.1145/1749603.1749606
- Hua, X. S., Lu, L., & Zhang, H. J., (2004). Optimization-based automated home video editing system. *IEEE Transactions on Multimedia*, 6(2), 190-202.
- Kress, G., & van Leeuwen, T., (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold Publishers.



- Lin, J. C., Wei, W. L., Yang, J., & Wang, H. M., (2017). Automatic music video generation based on simultaneous soundtrack recommendation and video editing. *Proceedings of the 25th ACM International Conference on Multimedia*, 1703-1711. DOI:10.1145/3123266.3123399
- Rodaway, P., (1994). *Sensuous geographies: Body, sense, and place*. London: Routledge.
- Voci, P., (2010). *China on video: Smaller-screen realities*. Routledge.
- Wang, J., (2024). Distinctive sound design: A comparative study on audiovisual analyses of different types of “immersive” short videos. *Journal of Social Science and Humanities Research*, 7(1), 45-58. DOI:10.53469/jsshl.2024.07(01).06
- Wang, Y., Liu, Z., & Huang, J. C., (2000). Multimedia content analysis: Using both audio and visual clues. *IEEE International Conference on Multimedia and Expo*.
- Zhang, H. M., (2024). Multimedia technologies for presenting poetry in online educational blogs: Interpreting the poems of Chinese poets in contemporary music of China. *Education and Information Technologies*, 29, 1150-1168.
- Zhang, W., (2021). *Viral Sounds: Music, Affect, Healing and Nostalgia in Wuhan*. Doctoral dissertation, The Chinese University of Hong Kong (Hong Kong) ProQuest Dissertations & Theses.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).