Constructing Zhaoqing's International Image through Visual and Auditory Signs

Han Xiao

Zhaoqing University, School of Foreign Studies, Zhaoqing, China

Abstract: This study takes the well-known city promotion film of Zhaoqing as its research subject. Drawing on French semiotician Roland Barthes' theory of signification, the paper analyzes the signs employed in the film and how they work together to construct and convey Zhaoqing's international image. Furthermore, the research attempts to explore the psychological impact of the film on viewers from a semiotic perspective. The findings indicate that the promotional film primarily utilizes visual signs such as characters, natural landscapes, and subtitles, along with auditory elements like background music and voiceover. The film not only effectively boosts Zhaoqing's recognition but also successfully meets the audience's psychological needs, including information seeking, curiosity, and entertainment.

Keywords: Zhaoqing, city image, international image, visual signs

I.   Introduction

With the rapid advancement of information technology, mass media have fundamentally transformed how people acquire knowledge and engage in communication. In particular, younger generations increasingly rely on visual resources to explore the world, share ideas, and seek entertainment. In this context, city promotion films have become a crucial tool for shaping urban image and enhancing a city's reputation. Zhaoqing is an important city in the Guangdong-Hong Kong-Macao Greater Bay Area, serving both as a strategic gateway connecting the Bay Area with southwest China and as its most beautiful ecological oasis. With profound Lingnan cultural heritage and an excellent ecological environment, the city is vigorously developing advanced manufacturing industries such as new energy vehicles and electronic information. It is actively building itself into a manufacturing hub, a health and leisure tourism destination, and a base for scientific and technological achievement transformation within the Bay Area, making it a strategic hinterland with tremendous development potential. Therefore, it is of considerable research importance to examine how city promotion films contribute to the successful construction of Zhaoqing's image and to analyze the mechanisms through which they influence audience psychology.

Against the backdrop, the study intends to probe into the following questions:

Which signs are involved in the Zhaoqing promotion film?

What kind of images are created through the promotional film?

How have the psychological needs of the audience been satisfied?

II.   Literature Review

The city image is a comprehensive representation of a city's intrinsic historical culture, developmental level, spiritual ethos, and its external landscape environment. It reflects the public's overall perception and emotional evaluation of a city. Against the backdrop of increasingly fierce globalization and regional competition, actively shaping a city's image is not only a strategic choice but also an inevitable requirement for urban development. In

an era of information overload, public perception of cities tends to be "labeled" and "symbolized." Passively accepting fragmented information may lead to one-sided or even negative impressions, making proactive image shaping a necessary measure to guide perception and master narrative initiative. Many cities face the planning dilemma of "a thousand cities with the same appearance," where historical context is fragmented and regional distinctiveness is blurred. By tapping into unique cultural genes and ecological endowments to create a differentiated image, cities can find a core pathway to break homogeneity and regain recognizability.

Some scholars employ big data from social media, text mining, deep learning, and sentiment analysis to examine the multidimensional and dynamic public perception of city images, moving beyond traditional survey methods to capture the complexity and real-time evolution of urban representation in the digital era. For example, Bai et al. (2024) analyzes city image through the lens of social media big data, utilizing Nanjing as a case to construct a perception model based on multi-source data and deep learning. It maps hotspot distribution, identifies perception characteristics such as "heritage monuments," "long history," and "culture," and evaluates the influence of functional, traffic, emotional, and visual perceptions on city image formation. Qu et al. (2024) leverages social media data as a novel lens for understanding and analyzing city image in the digital era. By collecting 370,000 Sina Weibo posts related to Guangzhou from 2019 to 2023, it employs text mining techniques, LDA topic modeling, and sentiment analysis to systematically examine how the city is perceived online. Jeong (2025) employs social media geographic data to analyze the dynamic and multi-themed nature of city images, moving beyond traditional survey methods. The study reveals the evolving city images of three South Korean destinations, such as Gongju's shift from a historical focus to a multifaceted identity and Jeju's transformation into a modern leisure hub, thereby validating and enriching knowledge on how city images change over time.

Some scholars research on the intersection of city image with specific domains. Zheng, Luo and Lu (2024) explore the underexamined role of linguistic soundscape, an integral dimension of linguistic landscape, in shaping city image. Focusing on Shenzhen, one of China's wealthiest cities, the research applies the City Identity System (CIS) framework to analyze how public address announcements in key transit hubs (metro stations, airport, ferry terminals) influence urban perception. By examining language patterns, functional purposes, and voice gender in recorded announcements, the study reveals that linguistic soundscape generally contributes positively to Shenzhen's city image, despite certain imbalances between informational and symbolic functions. The findings underscore the potential of intentional sound design in public spaces as a strategic tool for reinforcing and enhancing urban identity. Zhou, Ho and Mieiro (2024) investigate the role of government-led tourism promotion in shaping city image and influencing visiting intentions, using Macao as a case. Ma (2021) examines the impact of city image on international students' choice of a study destination in East Asia, specifically within the context of a flagship university in Taipei. The findings confirm that city image is a significant factor alongside university reputation, with its influence varying based on the student' program type, study level, identity, and region of origin.

Others probe into the selective construction and impact of city image in specific contexts. Qian (2022) put that the city image of Wangfujing on Instagram is narrowly constructed as an exotic hub of traditional food culture, driven by users' performance as urban adventurers. This selective visual storytelling, emphasizing the food market, reinforces a biased online identity that obscures its cosmopolitan reality and shapes public perception through platform-based communication. Su, Jia and Yang (2024) examine the international communication challenges and strategic optimization of Beijing's city image in the digital and VUCA era. It identifies specific communication inefficiencies in high-/low-context cultural translation, offering targeted strategies for global narrative adaptation.

Nevertheless, existing research has paid little attention to the city image of Zhaoqing, let alone studies from a social semiotic perspective. This gap is particularly noteworthy given Zhaoqing's unique status as a historically significant cultural city within the Guangdong-Hong Kong-Macau Greater Bay Area, which encompasses both profound Lingnan cultural heritage and ongoing modern urban development. From a social semiotic standpoint, the construction and dissemination of Zhaoqing's city image involve complex interactions among multiple meaning systems—such as its natural landscape symbols (e.g., the Seven Star Crags and

Dinghu Mountain), historical cultural symbols (e.g., the Song Dynasty city wall and traditional Cantonese architecture), and contemporary economic and social symbols (e.g., its role in regional integration and eco-tourism). These symbolic resources are dynamically interpreted and reconstructed across various media platforms and public discourses. Therefore, a social semiotic analysis can effectively reveal how different actors employ specific visual, textual, and multimodal signs to shape, negotiate, and even contest the representation of Zhaoqing, thereby deepening our understanding of the mechanisms behind city image formation in the context of cultural heritage preservation and regional modernization.

III.   Multimodal discourse analysis

Social semiotics is defined as "a socially shaped and culturally given semiotic resource for making meaning" (Kress 2010: 79). Socio semiotically viewed, modes are stratified structures, formed from a material dimension relating to the human sensory channels at one end, and linked to the most abstract social and cultural systems at the other (Kress and Van Leeuwen 2001). In visual discourse, any semiotic mode has to be able to project the relations between the producer of a sign and the receiver of that sign (Kress and van Leeuwen 2006). Visual practices construct cultural embodiments and symbols that identify a certain culture and people and thereby structure a body of representational social discourse (Aiello and Parry 2019; Tzanelli 2006). Analysing these particular semiotic orchestrations requires methods to recognise their discursive significance, both separately and in combination (Kress, 2010; Pauwels, 2012). A key approach for reconstructing meanings in discourse in communication studies, political science, sociology and linguistics has been to treat them as media frames. In their basic form, these are seen as structured representations of knowledge, which become materialised in utterances and organise patterns of discourse (Borah, 2011; Scheufele, 1999; Tannen, 1993). For multimodal analysis, Kress (2010), p. 149 holds that a "frame defines the world to be engaged with; it excludes and it includes; and in doing that it shapes, presents the world according to the interest and the principles of those who frame". Frames in discourse are therefore often viewed as evolving from framing as "meaning work – the struggle over the production of mobilising and countermobilizing ideas and meanings", as Benford and Snow (2000, p. 613) put it, by connecting materially available and semiotically conventionalized resources in an argumentative manner.

In the context of visual semiotics, this conceptualization of framing becomes particularly potent, as visual communication operates through a dense, often immediate, synthesis of symbolic, iconic, and indexical signs. A visual frame, therefore, is not merely a compositional border but a semiotic boundary that actively constructs a realm of relevance. It dictates what is visually presented and what is omitted, what is centered and what is marginalized, thereby performing a critical ideological function. For instance, a promotional video for Zhaoqing might frame the city exclusively through lingering shots of the Seven Star Crags and ancient temple roofs, while systematically excluding images of industrial zones or everyday urban congestion. This selective inclusion/exclusion transforms the city from a complex, living entity into a curated artefact, a "cultural oasis" or "scenic retreat", aligning perfectly with specific tourism and cultural heritage agendas.

Extending Kress's view, visual framing is inherently a process of semiotic orchestration. It involves the deliberate arrangement of visual resources—color, line, perspective, focus, montage sequence—to guide interpretation and evoke specific affective and cognitive responses. A low-angle shot of the Song Dynasty city wall can frame it as monumental and imposing, invoking awe and historical reverence (a top-down power relation between the viewed object and the spectator). Conversely, a eye-level, steady-cam walk through a bustling morning tea house frames the tradition as intimate, accessible, and lived-in, fostering a sense of experiential belonging. The frame thus acts as a directive, channeling the viewer's gaze and structuring their pathway through the visual narrative.

Benford and Snow's notion of "meaning work" finds a vivid manifestation in the visual arena, where framing is a site of continuous struggle and negotiation. Official city propaganda might frame Zhaoqing's development through sweeping aerial views of new bridges and highways, employing clean lines, blue skies, and fast-paced editing to semioticize "progress," "connectivity," and "modern efficiency." This constitutes a mobilizing frame aimed at attracting investment and showcasing governance. In counter-mobilization,

independent documentaries or social media posts by local residents might employ a contrasting frame: close-up, handheld shots of demolition sites adjacent to historic neighborhoods, using grainy textures, muted colors, and prolonged silences. This frame connects the material resource (the crumbling wall) with the semiotic convention of "nostalgia" or "loss," challenging the official narrative of seamless progress and mobilizing discourse around cultural preservation and community rights. The visual frame becomes the very battleground for these competing meaning systems.

Furthermore, in today's digital ecology, visual frames are highly intertextual and remediated. A single image of Dinghu Mountain sourced from an official promo can be reframed by netizens through memes, collages, or critical captions on platforms like Weibo or Xiaohongshu. The original frame's intended meaning ("natural purity," "spiritual tranquility") can be subverted, parodied, or repurposed, illustrating the dynamic, non-deterministic life of a visual frame once it enters public circulation. This process underscores that framing is not a one-time act by a creator but an ongoing social practice of interpretation and re-contextualization.

IV. Results and Findings

4.1 Analysis of Visual Resources

According to Kress and van Leeuwen's theory of 'visual grammar', the representational meaning of images can be divided into narrative processes and conceptual processes.

4.1.1 Visual Representational Meaning

Based on the theory of visual grammar, the representational meaning of imagery in the promotional film *Listening to the Mountains and Rivers of Zhaoqing* can be deconstructed systematically through three core dimensions: Participants define the subjects and objects that carry meaning, process types reveal the mechanisms of dynamic interaction and symbolic encoding, while circumstances construct the spatiotemporal and atmospheric fields in which the narrative unfolds. These three elements are interdependent and together transform visual components into a layered system of meaning.

The film features two interconnected categories of participants. Human participants are portrayed as both cultural practitioners and experiential witnesses: craftsmen carving Duan inkstones (Frame 1), lion dancers performing in rhythmic unison (Frame 2), monks engaging in temple rituals (Frame3), and tourists immersed in contemplation (Frame 4). These figures serve as dynamic mediators between the viewer and the landscape. Non-human participants are elevated to symbolic prominence: the karst peaks of Seven Star Crags mirrored in still waters, the primeval forests and waterfalls of Dinghu Mountain, the migratory birds of Star Lake Wetland, and cultural icons like the Duan inkstone and dancing lions. Together, they form a visual lexicon representing Zhaoqing's natural and spiritual heritage.

Frame 1

Frame 2

Frame 3

Frame 4

Narrative processes dominate through vectors of action and reaction. Action processes are conveyed through deliberate movements: the lion dance's vigorous leaps, the craftsman's focused carving, and birds soaring across wetlands. These create a rhythm of cultural vitality and ecological harmony. Reaction processes are shown through gazing, listening, or touching,tourists gazing at landscapes, children watching lion dances with wonder, visitors touching ancient artifacts. Their attentive postures guide the audience's emotional engagement. Conceptual processes emerge through symbolic montage: the mist-shrouded forests of Dinghu Mountain paired with incense smoke from Qingyun Temple, or inkstone textures juxtaposed with flowing water, encoding themes of tranquility, craftsmanship, and timelessness.

The film's settings are meticulously constructed through ambient and spatial circumstances. Locational circumstances include iconic sites: the cliffs of Lingyang Gorge, the courtyard of Bao Gong Temple, the stone paths of Mei'an, and the waterways of the West River. These are not mere backdrops but active spaces where culture and nature converge. Atmospheric circumstances are crucial: golden-hour lighting over Seven Star Crags, morning mist rising from Star Lake, twilight glowing on Song Dynasty walls, and the soft haze enveloping Dinghu Mountain. These elements temporalize the visuals, evoking contemplation, serenity, and a sense of sacred time. Instrumental circumstances, such as carving tools, lion dance costumes, and temple bells, authenticate the cultural practices depicted.

4.1.2 Visual Interactive Meaning

Within the framework of visual grammar, the meaning of visual interaction is primarily constructed through two modes: conceptual representation and narrative representation. Conceptual representation focuses on presenting the universal attributes, classification, and essence of things, creating a static, eternal, and analysable state of "being". Narrative representation, conversely, constructs actions, events, and changes through dynamic vector relationships, guiding the audience into a "happening" world characterised by temporal sequences and interactive connections. The Zhaoqing city promotional film masterfully integrates these two representational modes. It not only showcases Zhaoqing's static landscapes and spiritual essence but also narrates a dynamic tale of the symbiotic interplay between nature, history, and contemporary life. This achieves a profound persuasion, transforming the viewer's experience from a mere "cognitive geography" into an immersive "journey of discovery".

The promotional film extensively employs conceptual representations, distilling and crystallising Zhaoqing's core characteristics into a series of recognisable visual symbols and categories. Within the imagery, the profound forests of Dinghu Mountain (Frame 5), the karst pinnacles of Qixingyan, and the expansive reaches of the West River are not presented as backdrops for specific events, but rather as exemplars of the very category of 'Zhaoqing's natural endowment'. These shots often employ symmetrical, balanced compositions with measured or even static camera movements, evoking classical landscape paintings to emphasise timelessness, stability, and intrinsic harmony. For instance, the aerial view of Xinghu Lake integrates the lake, rock formations, and cityscape into a single frame, creating a structured diagram of geography and ecology that visually conveys the city's 'lake-city symbiosis' layout. Similarly, close-ups of cultural heritage items like Duan inkstones belong to this conceptual representation. The camera focuses on the stone's texture, stripping away the specific production process to highlight their material beauty and craftsmanship as 'cultural symbols'. This mode of representation does not narrate 'who did what and when,' but rather answers 'what is Zhaoqing?', an entity that embodies the essence of heaven and earth, crystallising millennia of cultural heritage. It constructs for the viewer a cognitive framework transcending time, inviting contemplation and appreciation.

Frame 5   Frame 6

In contrast to the contemplative nature of conceptual representation, the promotional film infuses vitality and temporality through powerful narrative representation. The silhouette of a runner traversing the Song Dynasty city wall forms a distinct vector of motion, transforming historical ruins into a vibrant space of the present. Athletes from the National Water Sports Training Base powerfully slice their oars through the water, with each stroke cutting through the waves and every spray of droplets forming precise vectors of forward momentum, embodying the ultimate synergy and competitive spirit of professional training (Frame 7). Citizens gather joyfully around a shared meal in a lush waterfront park. Their leaning postures, raised glasses, and interlocking gazes form a series of warm, inclusive vectors of interaction (Frame 8). This convivial tableau, framed by the natural scenery, visually constructs a tangible sense of community belonging and shared happiness, eloquently portraying the vibrant social fabric and human-centered vitality that define this truly livable city. These dynamic sequences weave micro-narratives, 'People live, exercise, preserve traditions, and create here.' The most poignant narrative thread visualises the theme of 'hearing': the photographer's lens tracks birds in flight, the musician's notes merge with the wind's melody, the visitor gazes contemplatively at the waterfall. Within these scenes, the subjects' gaze (visual vectors) and perceptual actions intertwine with the natural landscape, constructing a complete narrative arc: "subject (human), action (listening, perceiving), object (mountains and rivers)" narrative arc. It recounts not static landscapes, but the story of human encounters with nature, dialogue with the mountains and rivers, and the spiritual resonance gained therefrom. This compellingly guides viewers towards identifying with and yearning to experience this process firsthand.

Frame 7   Frame 8

The brilliance of *Listening to the Mountains and Rivers of Zhaoqing* lies in how these two representations are not disjointed, but rather cyclically intertwined and mutually deepening. The film frequently establishes scenes through conceptual panoramas (such as the vast West River), swiftly cutting to narrative close-ups (like fishermen casting nets or vessels navigating the waters), before concluding with an even grander conceptual shot (such as a sunset over the river), creating a rhythm of 'scene, event, sublimation'. For instance, the transition from the static panoramic view of the Seven Star Caves (conceptual) to the curious gaze of children touching stalactites within a cave (narrative), then dissolving into the ancient contours of rock walls beneath the starry sky (conceptual), accomplishes within minutes a profound cycle from cognition to experience and back to essence. This fusion endows Zhaoqing's image with both profound, enduring foundations (conferred by conceptual representation) and vibrant, approachable vitality (infused by narrative representation). Ultimately, through visual grammar, the promotional film transforms Zhaoqing from a geographical term into an emotional entity that can be heard, interacted with, and inhabited. It not only informs viewers of Zhaoqing's landscapes but invites them into an unfolding narrative, come here, and you shall not merely witness scenery, but become the protagonist of a story resonating with millennia-old mountains and rivers, attuned to vibrant vitality.

4.1.3 Visual Compositional Meaning

In the promotional film *Listening to the Mountains and Rivers of Zhaoqing*, visual composition is by no means a haphazard arrangement of elements. Instead, it rigorously adheres to the principles of compositional meaning within visual grammar. Through meticulous spatial layout, visual emphasis, and framing, it constructs Zhaoqing's natural landscapes and cultural spirit into a visually cohesive whole, rich in layers and meaning. This guides the viewer on a profound journey from sensory experience to deeper comprehension.

The positioning within the composition carries explicit informational value, establishing the fundamental semantic structure of the frame. The promotional film frequently employs a vertical composition: the sky, distant mountains, and the vast expanse of Xinghu Lake often occupy the upper portion of the frame (Frame 9), symbolising the "ideal" and the "sublime", representing the majestic power of nature and a transcendent spiritual realm. Human activities, such as strolling through alleyways, carving beside inkstones, or racing dragon boats (Frame 10), are firmly anchored in the lower portion of the frame, representing the "real" and the "foundational". This signifies that Zhaoqing's culture and way of life are deeply rooted in this land. This composition clearly conveys the core concept of "nature as the soul, humanity as the foundation". Simultaneously, the principle of centre and periphery is skilfully employed: whether a solitary boat adrift in the river or the central pattern of a Duan inkstone in the artisan's hands, each element is positioned at the visual focal point, becoming the indisputable core of its narrative. The surrounding landscapes or diffused environments thus become supporting peripheries, effectively eliminating visual distractions and reinforcing the purity of the theme.

| Frame 9 | Frame 10 |

To precisely direct the viewer's attention within complex scenes, the film employs multiple techniques of visual prominence. The most direct is contrast in scale: abrupt transitions from sweeping aerial panoramas to close-ups of artisans wielding chisels create rhythmic shifts. This dramatic change in perspective not only establishes pacing but also refocuses the grand concept of "mountains and rivers" onto the concrete, minute "people" and "objects", emphasising that humans are the perceivers and creators of the spirit within the landscape. The interplay of light, shadow, and colour shoulders the responsibility of emotional resonance: a beam of Tyndall light piercing the mist over Dinghu Mountain transforms a sliver of dawn in the shadowy forest into an overwhelming visual focal point, symbolising hope and spirituality; Upon the deep black base of a Duan inkstone, a natural stone pattern glows brilliantly under side lighting. The high saturation of its texture instantly elevates it from a utilitarian object to a work of art, emphasising both nature's marvels and the wisdom of human discovery.

Framing connects or separates elements through visual means, serving as the cornerstone for constructing both the logical structure and poetic atmosphere of an image. The film extensively employs subtle framing techniques: landscapes and cities, nature and human culture, are often seamlessly linked through flowing mist, rippling water surfaces, or soft light transitions, with scarcely any rigid boundaries discernible. This visually articulates the symbiotic relationship where 'the city resides within the mountains and waters, and the mountains and waters dwell within the city.' Simultaneously, the framing-within-framing technique is employed with profound poeticism: the latticework of ancient buildings, the arches beneath bridges, and the interlacing branches of trees all become natural picture frames, "enclosing" distant landscapes or human activities within their confines. This not only enhances the formal beauty and layered composition but also subtly suggests a perspective of contemplation and discovery, the beauty we perceive in Zhaoqing is defined and interpreted by culture (architecture), time (ancient bridges), and nature (trees). Each frame appears as a meticulously distilled cultural motif.

Through the systematic application of visual composition, the Zhaoqing promotional film successfully transforms geographical space into a text that can be "read". It does not merely showcase landscapes; rather, through spatial order, guided focal points and visual rhetoric, it silently articulates Zhaoqing's natural philosophy, cultural depth and aesthetic sensibilities. Ultimately, guided by the imagery, viewers come to "hear" the profound and resonant cultural echoes behind the mountains and rivers.

4.2 Intersemiotic Complementarity: The Interaction of Visual and Other Modes
4.2.1 Verbal-Visual Synergy: The Role of Language in Anchoring and Elaborating

Language and visuals form a finely tuned multimodal complementary system. Visual imagery delivers immediate sensory engagement and aesthetic impact, whilst language, encompassing narration, subtitles, and captions, collaborates with the visual through two core functions: anchoring and extension. Together, they construct layers of meaning richer and more profound than any single modality could achieve alone. This elevates the promotional film beyond mere scenic display, transforming it into a perceptible, interpretable, and resonant cultural narrative.

The anchoring function of language provides explicit interpretative pathways and semantic frameworks for open-ended visual imagery. When the lens captures the misty haze of Dinghu Mountain, the tranquil dawn light over Star Lake, or the craftsman's hands intently carving a Duan inkstone, these aesthetically rich scenes inherently possess multiple layers of meaning. At such moments, narration such as 'The mountains and waters have a clear sound' or subtitles like 'Moulding the heart with hands, preserving the craft for millennia' appear at opportune moments, anchoring the audience's sensory perceptions to specific cultural concepts such as 'spirituality,' 'artisan spirit,' and 'heritage.' Language here functions as a 'guiding star' of meaning, ensuring that open visual symbols are directed towards predetermined cultural interpretations. This prevents semantic dispersion and reinforces the core positioning of 'Zhaoqing, where nature and culture converge'.

Meanwhile, the expansive function of language profoundly broadens the temporal and spatial dimensions of visual imagery, pushing the boundaries of imagination. It serves as a bridge, connecting the "here and now" within the frame to the "river of history" and the "spiritual realm" beyond it. The phrase 'Millennia-old inkstone capital, ancient prefecture of Lingnan' transforms a static close-up of a Duan inkstone or a section of ancient city wall into an epic chronicle of profound civilisation. Similarly, the narration 'The waters of the West River, the lifeblood of Zhaoqing' symbolically links a close-up of a single water droplet to the grand, regional pulse of existence. Most ingeniously conceived is the title itself, Hearing the Mountains and Rivers. It actively invokes synesthesia, guiding viewers beyond mere 'viewing' to 'listen' to the stories behind the landscapes, the echoes of history, and the pulse of the city. This achieves a profoundly immersive experience, transcending visual perception to engage multiple senses in synaesthetic resonance.

Ultimately, this seamless synergy between language and visuals generates a synergistic communicative effect where the whole is greater than the sum of its parts. Visuals establish atmosphere, evoke emotion, and furnish compelling "evidence"; language structures logic, articulates themes, and elevates value. Both mutually reinforce and amplify each other. When audiences behold serene scenes of humanity coexisting harmoniously with nature, while simultaneously hearing the interpretation "A city of mountains and rivers, also a city of life", the fusion of sensory experience and rational understanding becomes complete. Together, they sculpt a multidimensional, richly textured image of Zhaoqing: not merely a picturesque tourist destination, but an ideal place to live, imbued with historical depth, cultural warmth and a philosophy of life. This achieves a profound transformation in the audience's mind, moving them from mere information reception to genuine value recognition. The deeper value of this collaborative relationship lies in its creation of a dialogic space for meaning generation. Language does not unilaterally interpret the imagery but engages in a nuanced interplay of 'question and answer' and 'complementarity' with the visuals. For instance, when the frame captures the dazzling lights of a modern metropolis, the narration might shift to recount the history of 'ancient Danzhou'. This juxtaposition of the visual "present" and the verbal 'past' creates a tension of temporal dialogue, suggesting that the city's vitality springs from its deep-rooted heritage. Conversely, when the frame focuses on the measured pace of traditional craftsmanship, the narration may employ modernist discourse about 'connecting the world,' endowing tradition with contemporary relevance. This strategy prevents the promotional film from becoming a one-dimensional hymn of praise, instead presenting a dynamic, dialectical image of the city. Simultaneously, the linguistic choices themselves, such as employing elegant, poetic vocabulary rather than straightforward explanatory text, align seamlessly with the film's overarching visual aesthetics (impressionistic, minimalist, refined), collectively establishing the promotional film's sophisticated cultural taste as a primary

message. It signals to viewers that this is not merely a tourism pitch, but an invitation to cultural appreciation. Thus, the synergy between language and visuals ultimately achieves more than merely conveying information about Zhaoqing; it successfully shapes and reinforces its cultural identity and brand ethos.

4.2.2 Sonic-Visual Orchestration: Music and Sound as Emotional and Temporal Framing

In the promotional film Hearing Zhaoqing's Mountains and Rivers, sound is not a mere appendage to the visuals but forms a meticulously orchestrated symphonic synergy with the imagery. Music, natural acoustics and human-made soundscapes interplay to establish a dual framework for the visual narrative: an emotional foundation and a temporal rhythm. This transforms the theme of 'hearing' from the film's title into an immersive experience that resonates with the audience.

The music establishes the overall emotional narrative arc. The soundtrack for promotional films typically opens with ethereal, melodious strings or piano, complementing sweeping natural vistas to cultivate an initial mood of tranquillity, expansiveness and a hint of sacredness. As the footage transitions to human-centred scenes (such as Duan inkstone carving or street life), the music often incorporates traditional Chinese instruments (like the guzheng or xiao), with melodies becoming more mellow and fluid, symbolising an emotional shift from reverence for nature to humanistic concern. The emotional dynamics of the music remain highly synchronised with the visual narrative. During climactic segments, such as the sunrise piercing the clouds or dragon boat races, it propels the mood to heightened intensity. In tranquil moments, like artisans deep in contemplation or rain falling upon Star Lake, it yields space, retreating to allow natural soundscapes to take precedence. Together, they weave an emotionally resonant arc that rises and falls with deliberate rhythm. The richly layered sound design constructs a multidimensional "soundscape space", greatly enhancing the realism and immersion of the visuals. The film meticulously records and blends multiple sound layers: the first layer comprises ambient background sounds such as wind rustling through trees, forest murmurs, and the gentle flow of water. These sustained low-frequency sounds establish the acoustic environment for each scene, allowing the visuals to "breathe". The second layer comprises signature action sounds, the scraping of carving tools against stone, paddles slicing through water, the crisp plop of tea leaves into a lidded bowl. These tactile, highly recognisable sounds transform visual close-ups into palpable auditory experiences, amplifying the intricate charm of craftsmanship and daily life. The third layer comprises intermittent humanistic soundscapes: the distant toll of bells, murmurs of local dialects, and children's laughter. These faint, intermittent sounds imbue the space with vitality and narrative, hinting at a living, vibrant Zhaoqing brimming with everyday life. Sound plays a pivotal role in shaping the temporal dimension. The tempo, rhythm, and editing points of the music correspond precisely to the visual transitions. Rapidly edited landscape montages, paired with upbeat musical passages, create a dynamic visual synthesis; whereas during extended shots contemplating landscapes or characters, the music slows, stretches, or yields entirely to ambient sounds, generating a 'suspended sense of time' that guides the audience towards profound contemplation and inner resonance. This synchronisation of audio-visual rhythms not only governs the film's narrative tempo but also mirrors the passage of the viewer's psychological time, at times exhilarating, at others contemplative.

The synergy of sound and imagery creates a 'synaesthetic realm' transcending mere summation. When the screen reveals misty landscapes reminiscent of ink wash paintings, ethereal music intertwines with reverberated bird calls and flowing water, collectively evoking the poetic spirit of 'paintings within landscapes'. When showcasing the intricate grain of Duan inkstones, deep, textured ambient sounds and solemn music reinforce the palpable weight of history. This synergy enables viewers not merely to 'see' the landscape, but to 'hear' its very essence and spirit, elevating the experience from sensory perception to a holistic emotional immersion. Sound thus becomes the key code for deciphering Zhaoqing's intrinsic character, working in tandem with the visual to fully construct a poetic world that invites dwelling, exploration, contemplation, and sensory engagement.

V.　　Conclusion

This study employs Visual Grammar Theory (Kress & van Leeuwen 2006) to conduct a multimodal

discourse analysis of the Zhaoqing city promotional film Return of the Auspicious Cranes, exploring how visual, auditory and verbal modalities synergistically construct the city's image and facilitate meaning integration.

Analysis reveals that through multi-layered modal strategies, the film successfully crafts a three-dimensional, compelling and resonant city image: on the historical dimension, classical musical accompaniment, archival imagery and rich visual textures underscore its profound historical heritage; on the cultural dimension, visual focus on symbols like Duan inkstones and guqin zithers, coupled with contextual narration, reinforces its cultural distinctiveness; On the landscape dimension, aerial panoramas, natural soundscapes, and poetic narration present the picturesque scenery of this 'city of landscape poetry'; on the developmental dimension, modern architecture, technological scenes, and an uplifting rhythm convey an innovative and enterprising urban spirit; on the lifestyle dimension, the fusion of everyday bustle, smiling faces, and ambient sounds showcases the city's vitality and human warmth; On an emotional level, immersive storytelling, lyrical music and evocative text ultimately stir viewers' emotional resonance and spiritual yearning for 'poetic dwelling'.

These multiple modalities are not merely layered but highly coordinated in rhythm, focus, and emotion: visual composition guides the gaze, music sets the mood, and narration provides anchors of meaning. The three elements mutually reinforce and resonate, collectively constructing a complete and compelling field of meaning. The promotional film thus not only conveys information effectively but profoundly engages viewers emotionally and fosters value recognition, successfully transforming the city's image into a tangible, contemplative, and aspirational mental landscape.

This study demonstrates that a meticulously crafted city promotional film constitutes a potent multimodal discourse practice. Through organic interaction between modalities, it achieves an integrated narrative of the city's history, culture, landscape, and spirit within a limited timeframe. Consequently, it facilitates effective communication of the city brand and identity construction at both cognitive and emotional levels within the audience.

References

[1] Aiello G & Katy P. 2019. *Visual communication: Understanding images in media culture*. London: SAGE Publications Limited.

[2] Bai X F et al. 2024. *Public perception of city image hotspots based on social media: A case study of Nanjing*, China. *IEEE Access*. 12: 116678-116692.

[3] Benford, R., Snow, D., 2000. Framing processes and social movements. *Annu. Rev. Sociol*. 26, 611–639.

[4] Borah, P., 2011. Conceptual issues in framing: a systematic examination of a decade's literature. *Journal of Communication*. 61 (2), 246-263.

[5]   Jeong H et al. 2025. An analysis of city image by exploiting social media: Toward a deeper understanding of multiple characteristics and their temporal changes using machine learning. *IEEE Transactions on Computational Social Systems*.

[6]   Kress, G. 2010. *Multimodality: A social semiotic approach to contemporary communication*. Abingdon: Routledge.

[7]   Kress G & Theo Van L. 2001. *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.

[8]   Kress G & Theo van L. 2006. *Reading images: The grammar of visual design*, 2nd edn. London: Routledge.

[9]   Ma A S H. 2022. Assessing the effects of university reputation and city image on international student destination choice: Evidence from a Flagship University in Taipei. *Education and Urban Society*. 54(8): 992-1009.

[10]   Scheufele, D.A., 1999. Framing as a theory of media effects. J. *Commun*. 49 (1),103–122.

[11]   Qian X. 2022. Three narrative patterns of the city image visually presented on Instagram under the influence of self-presentation. *Media, Culture & Society*. 44(6): 1149-1165.

[12]   Qu H M et al. 2024. Analysis of Guangzhou city image perception based on weibo text data (2019-2023). *Heliyon*. 10(7).

[13]   Su F, Jia, B X & Yang Y. 2024. Navigating challenges and psychological engines: case from international communication of Beijing city image. *International Communication of Chinese Culture*. 12(1): 1-16.

[14]   Tannen, D. (Ed.), 1993. *Framing in Discourse*. Oxford University Press, Oxford.

[15]   Tzanelli, R. 2006. Reel western fantasies: Portrait of a tourist imagination in The Beach. *Mobilities*. 1(1). 121–142.

[16]   Zhou J Q, Ho H W & Mieiro S. 2024. Government promotion and city image: visitor intentions in Macao. *International Journal of Tourism Cities*. 10 (3): 1124-1144.

[17]   Zheng R L, Luo J & Lu S G. 2024. Examining city image construction through the lens of linguistic soundscape: a case analysis of Shenzhen. *Landscape Research*. 49 (4): 554-567.