

An Integrated Study on Narrative Strategies and Effects of English Short Videos on Dongguan City Image from the Perspective of Multimodal Discourse Theory

Lifang Chen¹

¹*School of International Education, Dongguan City College, Dongguan, China*

ABSTRACT

In the context of globalization, urban image has become the core soft power for cities to participate in international competition. Short videos, with their audio-visual integration and lightweight dissemination characteristics, have gradually emerged as a crucial medium for cross-cultural communication of urban images. As an international lingua franca, English serves as a key medium for cities to break through cultural barriers and reach global audiences. Dongguan, as China's rapidly rising "world factory" since the reform and opening-up, has been transitioning from a "manufacturing hub" to a "manufacturing powerhouse" in recent years, generating new development momentum in areas such as technological innovation, cultural integration, and ecological livability. However, external perceptions of its urban image still predominantly remain tied to traditional manufacturing labels, with issues like stereotypical image representation and thin narrative in international communication. In this context, English short videos provide an opportunity for Dongguan to reconstruct a multifaceted urban image. Therefore, this paper will explore how to achieve effective urban image narration through the coordinated combination of multimodal discourse (such as scene construction of visual symbols, linguistic expression and musical accompaniment of auditory symbols, and narrative logic of textual symbols).

Keywords: *Multimodal discourse theory, Urban image; Narrative strategies, Communication effects.*

1. INTRODUCTION: RESEARCH BACKGROUND AND SIGNIFICANCE

1.1 Research Background

Currently, short videos, with their fast-paced and concise dissemination characteristics and the expressive advantages of multi-symbol integration, have become the core medium for cross-cultural communication of urban images. As an important manufacturing base in China, Dongguan occupies a pivotal position in the global industrial landscape. However, international perceptions of it remain largely confined to the singular label of "world factory," failing to fully showcase its multifaceted features such as technological innovation, cultural heritage, and livability. How to reconstruct Dongguan's multidimensional image through English short videos as a cross-linguistic medium,

using narrative strategies that align with international audiences' reception habits, has become a crucial challenge in external communication practices.

Multimodal discourse theory, rooted in social semiotics, positions images, sounds, and text as a synergistic system for meaning production. While language conveys core messages, visual elements construct scenes, and audio components (background music, voiceovers, sound effects) shape emotional resonance—all working together to generate meaning. This framework aligns with the communication characteristics of English short videos: they require linguistic cross-cultural communication while relying on visual and auditory symbols to amplify impact. This provides a theoretical framework for analyzing the multimodal nature of short video content.

Current research predominantly focuses on the narrative content or dissemination channels of urban image short videos, with limited analysis from a multimodal integration perspective. Studies on English-language short videos mainly emphasize linguistic accuracy while inadequately exploring modal synergy effects and international audience reception mechanisms. Researches on Dongguan's urban image communication remains concentrated on traditional media or single-modality texts, showing relatively weak attention to this emerging medium. Against this backdrop, this study attempts to integrate multimodal discourse theory with communication effectiveness research, systematically analyzing the multimodal composition of narrative strategies in Dongguan's English-language short videos. It explores how different modal combinations influence international audiences' cognition and attitudes, ultimately proposing optimization suggestions. Through content analysis, this study identifies multimodal usage characteristics in Dongguan's English-language short videos—aerial shots showcasing Songshan Lake High-Tech Zone (visual), English voiceovers introducing industrial chains (text), and electronic music enhancing innovation perception (auditory). Cultural themes employ slow-motion techniques to present Qianjiao lamp craftsmanship (visual), Cantonese nursery rhymes to create ambiance (auditory), and English subtitles to interpret cultural connotations (text). Simultaneously, through questionnaires and in-depth interviews, the study measures international audiences' acceptance of different modal combinations: whether visual scene-based elements enhance innovative image recognition, emotional auditory elements strengthen cultural identity, and as if concise linguistic expression reduces comprehension barriers. Meanwhile, this study advances multimodal discourse theory from textual analysis to integrated research on communication effects, expanding its application scenarios in the field of cross-cultural communication of urban image.

1.2 Significance of This Study

This study addresses the shortcomings of existing research that overemphasizes modal descriptions while neglecting effect correlation, providing a comprehensive strategy-effect analytical framework for multimodal

communication studies. Practically, it offers strategic references for Dongguan and similar cities to enhance the effectiveness of English short video dissemination, promoting the multidimensional internationalization of Dongguan's city image. Additionally, the research on Dongguan provides valuable experience for other manufacturing powerhouses like Foshan and Huizhou in conducting English short video city image communication, assisting Chinese cities in building more distinctive and approachable images within the global discourse system.

2. LITERATURE REVIEW

In the era of deep integration between globalization and media technology, urban image has evolved from a "local governance objective" to an "international communication node" — It not only serves as an external manifestation of a city's comprehensive strength (Ruan & Feng, 2007), but also becomes a micro-carrier of national identity (Guo, 2018). The rise of short videos provides a new pathway for cross-cultural communication of urban image: its "short, concise, and fast" format aligns with contemporary audiences' attention patterns (Xiao & Zhang, 2020), while its "visual-first, multimodal synergy" characteristic transcends linguistic barriers to directly convey a city's culture and identity (Yan, 2023). As English is the global lingua franca, short video narratives have become a key channel for cities to connect with international audiences — Overseas publishers utilize platforms like YouTube to "tell China's urban stories well" through English and multimodal symbol combinations (Wang, 2024).

Multimodal discourse theory provides a methodological framework for analyzing this complex process. Originating from Systemic Functional Linguistics (Halliday, 1978) and expanded by Kress & van Leeuwen's visual grammar (1996), it emphasizes the collaborative construction of meaning through multiple symbolic modalities including language, images, sound, and interaction. Chinese scholar Xiao Jun (2017) further introduced this theory into new media cross-cultural communication research, noting that multimodality can "achieve shared meaning across cultural contexts through modal elements like emojis and short videos." However, existing studies have yet to adequately address three practical

challenges: First, how to adapt modality combinations in English short videos to align with overseas audiences' preferences. Second, which strategies can effectively enhance cognitive, emotional, and behavioral effects when integrating narrative strategies with communication outcomes. Third, how can cities undergoing manufacturing transformation—such as Dongguan, the world factory effectively convey their new image as smart manufacturing hubs through multimodal narratives. These questions form the foundation of this study.

Existing research can be categorized into three main themes: the application of multimodal discourse theory in media communication, multimodal narrative strategies for urban image promotion, and the dissemination effects and influencing factors of short video-based city branding. These three dimensions progressively unfold from theoretical applications to strategic analysis, and finally to effectiveness validation, collectively forming a comprehensive research framework for multimodal urban image communication.

2.1 Applications of Multimodal Discourse Theory in Media

The core hypothesis of multimodal discourse theory posits that synergistic symbol-modes can achieve more complete meaning construction (Kress & van Leeuwen, 1996). Current research has applied this framework to various media types including tourism promotion, news reporting, and education, demonstrating the hypothesis's universal applicability.

In tourism promotion studies, Yanwei & Hassan (2023) analyzed the multimodal synergy between text and imagery through Beijing's tourism promotional videos. Shi (2024), examining Hangzhou's Asian Games promotional videos, applied visual grammar to analyze the roles of text, visuals, and sound. This tripartite approach collectively shaped the city's image as "vibrant, open, and inclusive."

2.2 Researches on Multimodal Narrative Strategies of City Image

The core of this theme is how to convey the core characteristics of a city through multimodal

symbol combination. Existing researches focus on the correspondence between symbol and meaning, forming two representative views: cultural core theory and audience adaptation theory.

The "Cultural Core Theory" posits that multimodal narratives should revolve around a city's core cultural characteristics, reinforcing cultural identity through modal synergy. He Jingping (2018) analyzed the multimodal discourse in the city image video advertisements of Quanzhou, a node city under the Belt and Road Initiative, using linguistic texts to clarify cultural positioning, images to present the connection between history and modernity, and sounds to strengthen regional cultural identity. These three elements collectively construct a city image of cultural inheritance and openness. Liu Lizhi (2023) took Guangzhou as an example, proposing that urban lifestyle is the narrative core. Through visual symbols like "morning tea", "arcades", and "Cantonese opera", combined with the English narration Cantonese life is slow and flavorful, it conveys a city culture of leisure and diversity.

The "audience adaptation theory" emphasizes that multimodal narratives must align with the cultural context and media consumption habits of their audience. Wang Yanqi (2024) studied Chongqing's city image videos on YouTube, revealing that overseas creators frequently employ high-visibility visual elements like "hot pot", "Hongyadong," and "Liziba Light Rail," paired with "simple English narration" to position the city as "distinctive and livable." — This strategy caters to international audiences' preference for novelty and accessibility. Yan Yan (2023), using Changsha as a case study, identified issues in short video multimodal narratives.

2.3 Researches on the Communication Effect and Influencing Factors of Short Videos of City Image

This part focuses on how multimodal narrative affects audience cognition. Existing studies are carried out from the two aspects of quantitative measurement and qualitative analysis, forming two types of views: effect dimension theory and factor mediation theory.

The "Effect Dimension Theory" categorizes communication effects into three dimensions:

cognitive, emotional, and behavioral, emphasizing the differential impacts of multimodal narratives across these dimensions. Wang Yanqi (2024) used Python to scrape comment data from Chongqing city image videos on YouTube, revealing that "the uniqueness of visual symbols significantly enhanced cognitive effects, while the comprehensibility of English narration boosted emotional engagement. Interactive elements like Subscribe to my channel for more China videos facilitated behavioral effects such as subscriptions and sharing. Yan Yan (2023), using Changsha as a case study, demonstrated that multi-stakeholder communication as citizens' authenticity resonated with audiences. Guo Ke et al. (2018) analyzed international media coverage of Shanghai, finding that multimodal reporting improved cognitive effects by 30% compared to text-only reports. Notably, combinations of technological innovation and cultural events (e.g., Shanghai International Film Festival visuals) most effectively reinforced behavioral effects like investment willingness.

The mediation theory examines mediating variables affecting communication effectiveness, such as "cultural distance" and "audience media consumption patterns." Zahra (2021) studied image dissemination among Syrian refugees on Facebook and found that "cultural distance" serves as a key mediator— For Western audiences, "daily life videos of refugees" evoke stronger sympathy than "static images" because dynamic visuals "reduce cultural barriers."

Overall, existing research has provided a theoretical foundation for the integration of multimodal narratives and city image communication. However, studies on English short videos in China's transitional cities suffer from limitations such as fragmented strategic analysis, weakened effect relevance, and absence of regional characteristics. As a city combining industrial transformation typicality with cultural diversity, Dongguan's multimodal narratives in English short videos need to both respond to the global perception of China Manufacturing and convey the new connotation of Quality Dongguan. Current research has not systematically addressed this issue, leaving room for this study to explore.

Based on this, this study employs multimodal discourse theory to investigate the synergistic

mechanisms between visual and textual symbols in English short videos, revealing how these elements collectively shape Dongguan's multidimensional urban identity as China's Advanced Manufacturing Hub, Cultural Gateway of Lingnan, and Quality Living Destination. The research identifies three core narrative strategies: 1) Transforming abstract city positioning into relatable life scenarios through contextual storytelling; 2) Conveying urban ethos via character-driven narratives that mirror individual destinies; 3) Optimizing symbolic presentation through cross-cultural adaptation to minimize communication barriers. Additionally, an evaluation framework is developed from cognitive, emotional, and behavioral perspectives to assess how these narrative approaches impact overseas audiences' perceptions and behaviors.

3. A BRIEF REVIEW OF MULTIMODAL DISCOURSE THEORY

Multimodal discourse theory originated from systemic functional linguistics breakthrough in defining language boundaries, expanding discourse analysis from single textual symbols to the interactive relationships among visual, auditory, tactile, and other multimodal symbolic resources. Its core logic posits that meaning production does not rely on isolated symbol systems but is achieved through the coordination, complementarity, or conflict of multiple modes—. For instance, Kress and Van Leeuwen's visual grammar deconstructs images into three dimensions: representation, interaction, and composition. When Halliday's meta-functional theory extends to multimodal contexts, it requires considering how each mode fulfills conceptual functions (conveying factual information), interpersonal functions (establishing communication relationships), and discourse functions (constructing narrative logic). This theory provides foundational logic for analyzing hyper-textual communication forms like short videos: As carriers of Dongguan's city image, English short videos essentially form a "network of meanings" through multimodal elements including imagery (landmark buildings, industrial scenes, cultural activities), sound (English narration, Lingnan music, environmental effects), text (subtitles, copywriting), and cinematography (shot sizes, movement, editing). Their narrative strategies are realized through the selection and configuration of these

modal resources, while communication effectiveness depends on their meaning negotiation within cross-cultural contexts.

The construction of a multimodal discourse analysis framework must center on "meaning production," transforming the core elements of Dongguan's city image into communicable modality symbols. This process achieves narrative objectives through strategic allocation of modality resources, then optimizes modality selection based on feedback from communication effectiveness—serving both as a practical application of multimodal discourse theory and a logical closed-loop of from symbol to meaning in cross-cultural communication of urban images. Grounded in social semiotics, multimodal discourse theory defines modality as a symbolic system that conveys meaning through specific sensory channels, including visual (images, colors, compositions, camera movements), auditory (speech, music, sound effects), and textual (written content, subtitles, titles). These constitute the core symbolic carriers for English short videos in shaping Dongguan's city image. The complementarity between modality types forms one of its core logics, specifically manifested through temporal synchronization, spatial resonance, and meaning reinforcement. Additionally, the functional differentiation of modality determines their roles in storytelling: representational functions directly showcase industrial characteristics through visual modality and textual modality; interactive functions enhance audience immersion via auditory modality tones and visual modality angles; while constitutive functions structure narratives through modality arrangements. More importantly, multimodal meaning generation is always embedded in sociocultural contexts.

At its core function, multimodal discourse theory transforms abstract urban imagery into perceptible, comprehensible narratives through coordinated symbolic systems. This approach proves crucial for English short videos promoting Dongguan's city image. By combining visual immersive visuals, auditory engaging experiences, and textual precision, these elements collectively create a three-dimensional, vibrant, and cross-cultural compatible representation of Dongguan.

4. ANALYSIS OF MULTIMODAL NARRATIVE STRATEGIES IN ENGLISH SHORT VIDEOS OF DONGGUAN

Multimodal discourse theory conceptualizes language, visual, and auditory systems as a collaborative network for meaning construction. The narrative strategy of Dongguan's urban image English short videos transforms abstract city characteristics into perceptible communication texts through these modal interactions. Visual symbols, serving as the core narrative medium, employ montage contrast to interweave the city's past and present: juxtaposing red brick factory walls with glass facades of Songshan Lake smart factories, contrasting crowded stalls at Humen's old fashion wholesale market with runway stages of modern fashion weeks. These concrete spatial transformations narrate the story of transitioning from manufacturing to innovation. Cinematic language favors micro-perspectives—the steaming shrimp dumpling buns in breakfast shops, elderly artisans molding malt sugar figurines at alley entrances, and the silhouettes of teenagers leaping on basketball courts. These everyday scenes convey urban warmth through small narratives, avoiding the alienation inherent in grandiose storytelling.

The auditory modality serves dual functions of emotional resonance and cultural translation. English dubbing abandons didactic tones in favor of a conversational approach, such as opening narratives with phrases like Imagine waking up to the aroma of dim sum and the hum of robots, using relatable imagery to connect with audiences. Background music often incorporates elements of Lingnan folk music: the layered blend of Cantonese opera's gaohu melodies with electronic rhythms, or the interplay between wooden mallets during incense-making and lively piano melodies in cafes. This approach preserves cultural distinctiveness while aligning with international audiences' aesthetic preferences. The integration of urban sounds enhances authenticity—From street vendors' cries on old alleys to factory machinery noises and basketball arena cheers, these "non-verbal symbols" materialize the city's vibrant atmosphere and innovative spirit, transforming auditory elements into an emotional bridge between audiences and their urban environment.

The textual modality serves as a meaning anchor for visual and auditory elements, transforming abstract concepts into relatable narratives through humanized storytelling. For instance, when describing the transformation of manufacturing, instead of stating Dongguan achieved industrial upgrading, it portrays "Mr.Chen, a third-generation factory owner, transformed his father's textile mill into an AI-driven hub where robots assemble smartphones and design custom clothing," using a specific individual's experience to convey the city's developmental trajectory. The repetition of keywords reinforces the core themes— alternating phrases like Made in Dongguan and Created in Dongguan, while parallel sentences pairing Tradition with Innovation weave through the narrative, making the text a cohesive thread that connects multimodal elements. In addition, the text often forms a dialogue with the visual: when the camera focuses on the carved beams of the ancestral hall in Nanshe Ancient Village, the text adds that These wooden carvings are more than 500 years old—today, they inspire local designers to blend ancient patterns with modern fashion, transforming the aesthetic value of traditional architecture into a source of contemporary creativity.

The synergistic interaction between media formats is the essence of narrative. Visual elements like Cantonese opera performers applying facial makeup are accompanied by synchronized sound effects of makeup brushes, while text explains Every brushstroke tells a story from Cantonese opera—an art form that has been ingrained in Dongguan's cultural DNA for centuries. In slow-motion basketball footage, audience cheers blend with the message: Basketball isn't just a sport here—it's a language connecting grandparents and grandchildren, elevating the sports scene into the city's spiritual symbol. This "1+1> 2" media integration not only preserves cultural authenticity but also achieves cross-cultural communication of meaning.

In terms of narrative structure, short videos predominantly adopt the question-answer or suspense-reveal format. They typically begin with a provocative question like Can a city renowned for factories also become an art hub? followed by responses through museum exhibitions, creative workshops in cultural parks, and design week artworks. The narrative culminates with the theme

statement: Dongguan doesn't choose between manufacturing and creativity—it merges them, using dramatic contrast to emphasize the city's uniqueness. In essence, the multimodal storytelling approach in Dongguan's English short videos for city branding transforms the city's hard power (industrial and technological strengths) and "soft power" (cultural heritage and lifestyle) into perceptible narratives through visual engagement, auditory immersion, and textual exploration, these videos enable audiences to not only understand Dongguan's identity but also experience its essence firsthand. This narrative strategy aligns with the core concept of synergistic meaning generation in multimodal discourse theory while addressing the principle of emotional connection priority in cross-cultural communication, offering a replicable model for global city promotion.

5. AN EMPIRICAL STUDY ON THE EFFECT OF ENGLISH SHORT VIDEOS ON THE IMAGE OF DONGGUAN CITY

5.1 Research Design and Data Collection:

As a pivotal hub in the Guangdong-Hong Kong-Macao Greater Bay Area, Dongguan's international branding requires adaptation to short video's multimodal nature— where visual, auditory, and textual elements synergistically construct meaning. The core mission of English-language short videos is to present a multidimensional image to global audiences: a hub for smart manufacturing, a renowned Lingnan cultural city, and a livable and business-friendly metropolis. This study employs a mixed-method research framework grounded in multimodal discourse theory, integrating qualitative narrative analysis with quantitative communication effectiveness measurement to achieve mutual validation between theoretical perspectives and empirical data.

The research focuses on the English short videos released by Dongguan City Image between 2024 and 2025, sourced from official channels (Dongguan Municipal Government's overseas social media accounts and the Dongguan Culture and Tourism Bureau's YouTube channel) and self-media platforms (overseas travel bloggers and industry observers 'TikTok/Instagram accounts). The sample selection adheres to three criteria:

content themes covering four major dimensions— industrial (Songshan Lake High-tech Zone, Huawei European Town, etc.), cultural (Cantonese opera performances, Liaobu Xiang Market, etc.), lifestyle (Wanyi cuisine, community governance, etc.), and ecological (Dongjiang Wetland, Tongsha Ecological Park, etc.)—to comprehensively reflect the city's multifaceted image; representative dissemination scope (over 10,000 plays); and balanced sample types (30 official and 30 self-media samples, totaling 60 analytical samples).

Multimodal discourse analysis, grounded in Chris Chomsky's Visual Grammar and Halliday's Systemic Functional Linguistics, constructs an encoding framework across four dimensions: 1) Linguistic modality, stylistic register, thematic keywords ; 2) Visual modality , camera movements, color schemes; 3) Auditory modality, sound effects , and 4) Interactive modality, interactive prompts. The process employs frame-by-frame coding with NVivo 12 software, followed by independent analysis by dual coders (Kappa coefficient ≥ 0.8) to ensure reliability, ultimately refining typical multimodal narrative patterns through thematic clustering.

The measurement of communication effects is conducted through three-dimensional indicators: breadth, depth, and validity. The breadth dimension extracts backend platform data (play counts, share counts, cross-platform reposts). The depth dimension assesses user engagement through metrics like like rate (likes/total plays), comment rate (comments/total plays), and collection rate (collections/total plays). The validity dimension combines objective data with subjective perception. Objectively, Python web crawlers are used to scrape comments, and VADER tool is applied to annotate sentiment tendencies (accuracy ≥ 0.85). Subjectively, questionnaires were distributed via Amazon MTurk platform to overseas audiences in North America, Europe, and Southeast Asia (500 valid responses). These questionnaires evaluate four dimensions: industrial perception ("Is Dongguan's manufacturing industry primarily labor-intensive or high-tech?"), cultural perception ("Does Dongguan culture lean more towards traditional Cantonese culture or modern urban culture?"), lifestyle perception (Is Dongguan a livable city or a crowded one?), and ecological perception (Is Dongguan's ecology predominantly green or polluted?). The questionnaire underwent

pre-research optimization (Cronbach's α coefficient ≥ 0.72) to ensure reliability and validity.

The research methodology follows a logical sequence of sample screening, synchronized data collection, multimodal coding, quantification of effect indicators, and analysis of correlation mechanisms: First, samples are screened through thematic and dissemination metrics to simultaneously extract multimodal texts and dissemination data. Next, narrative strategies are decoded using NVivo coding, followed by SPSS analysis of dissemination effect characteristics. Finally, regression analysis explores the correlation between multimodal narrative patterns (such as first-person blogger accounts + aerial industrial park shots + warm-toned visuals + natural sound effects) and dissemination outcomes (high engagement rates, positive reviews, and industry recognition shifting toward high-end manufacturing).

6. RESULTS AND FINDINGS

6.1 Dimensional Analysis of the Effect

In terms of Cognitive Perception: 68% of overseas respondents learned about Dongguan's Cantonese opera and GuanXiang incense through short videos for the first time, while 55% corrected their preconceived notion that Dongguan is just a factory hub. They highlighted its smart manufacturing and new energy industries, noting Dongguan's culture differs from Guangzhou and Shenzhen with its distinctive water town charm. However, regional preferences remain distinct: Western audiences focus more on smart manufacturing, Southeast Asian audiences show greater interest in "Lingnan culture", while African audiences express curiosity about life in Xin-Dongguan.

As for Emotional Perception: 82% of audiences held positive perceptions toward Dongguan, with cultural ambience being the most favored element (88%). Overseas viewers particularly appreciated the exquisite costumes in Cantonese opera performances and the vibrant dragon boat races, while stories of ordinary citizens garnered the highest recognition (79%)— They found these narratives authentic and relatable, mirroring entrepreneurs' experiences. Negative feedback mainly stemmed from audio-visual incongruity.

another cultural video's overly academic English narration (Guanxiang, a representative of intangible cultural heritage with significant historical value) drew complaints from 20% of viewers who preferred the simplified phrasing Guanxiang is a unique incense crafted in Dongguan for over 1,000 years.

Regarding the Behavioral Level: 51% of overseas respondents expressed willingness to recommend Dongguan to friends, while 38% indicated interest in visiting or conducting business research in Dongguan. Younger demographics (18-35 years old) demonstrated stronger willingness (62%), as they are more susceptible to being influenced by story-driven narratives and visual impact. Business professionals (36-50 years old) prioritized the accuracy of industry data, with one video mentioning 90% new energy vehicle component supply rate in Dongguan. After verifying this with official statistics from Dongguan Bureau of Statistics, respondents confirmed it as a worthy investment opportunity.

7. STUDY ON THE INTEGRATION BETWEEN NARRATIVE STRATEGY AND THE EFFECT

Within the framework of multimodal discourse theory, the narrative strategy of Dongguan's English short videos for city branding centers on the synergistic integration of visual, auditory, and textual elements. This synergy transcends mere element accumulation, forming a systematic synthesis around the city's core identity. From its transformation from being the "world's factory" to innovating smart manufacturing, to preserving the living heritage of Lingnan culture, from creating eco-friendly living environments to cultivating open foreign trade capabilities, each urban characteristic is conveyed through complementary multimodal symbols that collectively achieve precise communication outcomes. Fundamentally, the integration of multimodal strategies and their effects stems from the alignment between symbolic systems and audience cognition. As a major international trade hub, Dongguan's English videos primarily target overseas investors, business partners, and potential tourists whose cognitive patterns favor intuitive presentation and contextual relevance. The visual element combines time-lapse shots of the Hong Kong International Container

Terminal with English narration emphasizing its status as one of the world's top 20 container ports serving over 100 countries and regions, while auditory elements blend the metallic hum of port cranes with the rhythmic whistles of cargo ships. This visual-auditory synthesis transforms the city's "openness" into quantifiable symbols that directly address overseas audiences' core demand for business convenience. Conversely, scenes of morning jogging at Huangqishan Forest Park (visual), ambient sounds of birdsong and citizens' laughter (auditory), along with the English narration highlighting 92% forest coverage enabling harmonious coexistence between urban and natural environments, precisely meet international tourists' expectations for livability. Data feedback shows that the sharing rate of this kind of demand-oriented multimodal integrated video on overseas platforms is 40% higher than that of the general city scenery video, because the integration of narrative strategies closely follows the needs of the audience, and the communication effect is changed from "passive viewing" to "active communication".

In essence, the narrative strategy of Dongguan's city image English short videos integrates multimodal symbols to construct a perceptible, empathetic, and relatable to city identity. The enhanced communication effectiveness results from aligning this integrated approach with audience cognitive patterns. From rationally conveying smart manufacturing transformation to emotionally resonating with cultural heritage preservation, from presenting open inclusiveness scenarios to advocating ecological livability values, each multimodal integration strengthens the core identity of the city while precisely empowering communication outcomes. This integration not only demonstrates the practical implementation of multimodal discourse theory but also embodies Dongguan's core logic for globalizing urban image communication in the contemporary context.

8. OPTIMIZATION STRATEGIES FOR DISSEMINATING DONGGUAN'S CITY IMAGE

Based on the above discussion, it can be concluded that the international communication of English short videos of Dongguan's image requires a coordinated construction of linguistic, visual, and

auditory elements to create a perceptible urban narrative. On the linguistic level, rigid expressions should be avoided in favor of more approachable local perspectives. For instance, foreign-funded enterprise managers operating in Dongguan could narrate in English: Our supply chain helps European clients launch new products three months ahead of schedule, transforming abstract concepts like supply chain into concrete examples. Visually, distinctive symbols should be distilled: close-ups of robotic arms working in Songshan Lake Smart Factory, juxtapositions of bricks from Humen Fortress and high-speed trains at the station, and group shots of cheering fans at Dongguan Basketball Center. These elements are linked through contrast montage to connect history with modernity, manufacturing with vitality. On the auditory dimension, gentle melodies from Cantonese opera tunes or contrasts between factory machinery sounds and the aromatic ambiance of cultural and creative park cafes can be incorporated. This sonic enhancement strengthens scene memorability while reducing cultural comprehension barriers.

The narrative perspective must shift from grand achievements" to "micro-level individuals: Focus on Nigerian merchants 'decade-long journey from street vending to cross-border e-commerce ventures, connecting Dongguan's logistics efficiency with its inclusive atmosphere; document the process of a thousand-angle lantern inheritor training apprentices in adjusting lanterns, transforming intangible cultural heritage preservation into concrete scenes like "grandmothers teaching granddaughters to fold lantern corners", using personal stories to bridge the gap with overseas audiences. Content positioning should cater to segmented audience needs: For trade professionals, highlight the case of Chang 'an Town Precision Hardware Factory supplying Tesla, featuring the factory manager explaining in English We've improved precision from 0.1mm to 0.01mm; for culture enthusiasts, have foreign bloggers learn dumpling wrapping techniques from artisans, capturing the delicate details of rice-filled bamboo leaves with captions like "Each crease symbolizes hope for reunion", reducing cultural barriers through immersive experiences.

The feedback mechanism should be implemented throughout the entire process: By analyzing platform data to track audience behavior

—— If a Manufacturing Upgrade video achieves an 85% completion rate with frequent mentions of impressive supply chain in comments, we should deepen such content. If a Cultural Experience video receives three times the average number of likes and overseas users comment want to visit the lantern workshop, we should launch a Cultural Heritage Experience Route series. Additionally, we can directly collect user needs through overseas surveys and use feedback to calibrate narrative direction.

Furthermore, the interpretation of cultural symbols proves crucial. For the Thousand-Angle Lanterns, we first captured panoramic shots before focusing on artisans meticulously attaching paper decorations, accompanied by captions like 600 years of craftsmanship, each angle symbolizing a family's aspiration for prosperity. Ultimately, through multimodal integration, Dongguan's image crystallized into core labels: innovative, historic, vibrant. This transformation guides overseas audiences from unfamiliarity to familiarity and ultimately to recognition — showing not an abstract manufacturing city, but a warm, story-rich, and future-oriented living space. This integration fundamentally shifts city branding from propaganda to dialogue — replacing grand narratives with individual stories, conceptual indoctrination with scene-based perception, and one-way communication with two-way interaction— allowing Dongguan's image to truly resonate with global audiences.

9. CONCLUSION

The discussion above demonstrates that multimodal discourse theory provides a systematic analytical framework and optimization pathway for the narrative practices of English short videos promoting Dongguan's city image. The study reveals that the coordinated operation of visual, auditory, and textual modalities constitutes the core logic—— for constructing a three-dimensional urban image and enhancing communication effectiveness. Visual symbols concretize Dongguan's multifaceted identity through contextualized storytelling, auditory elements amplify emotional resonance via emotional rendering, while textual components refine core identity markers through "precise anchoring. This organic integration effectively overcomes the

limitations of single-modality information, establishing a multidimensional urban image that is visually engaging, auditorily immersive, and linguistically comprehensible. Consequently, this approach significantly deepens English audiences' cognitive understanding and emotional identification with Dongguan.

Empirical data reveals that the immersive experience and cross-cultural adaptability of multimodal narratives are key factors in enhancing communication effectiveness. On one hand, the immersive quality of multimodal content significantly increases viewers' engagement time and interaction rates. On the other hand, cross-cultural adaptation reduces cultural barriers. This approach preserves the authenticity of Dongguan's cultural characteristics while ensuring English-speaking audiences can understand them, effectively boosting communication efficacy.

In general, this study not only provides theoretical support and practical reference for the external communication of Dongguan city image, but also accumulates empirical cases for the application of multimodal discourse theory in the field of urban communication.

ACKNOWLEDGMENTS

This paper was funded by the Young Faculty Research Fund of Dongguan City College (Grant No. 2022QJY015R).

REFERENCES

- [1] Du Wei. The Cultural Tourism Integration Effect of Ice and Snow Elements in Urban Image Shaping—— Review of Urban Image Shaping and Cultural Element Expression—— From the Perspective of Cultural Tourism Integration. *Modern Urban Studies*, 2025(05).
- [2] Guo Ke, Chen Yue, Du Yan. The Formation Mechanism and Theoretical Interpretation of Global City Image Communication: A Case Study of Shanghai's Urban Image. *Journal of Journalism University*, 2018(06).
- [3] He Jingping. Multimodal Discourse Analysis of Urban Image Advertising under the Belt and Road Initiative. *Journal of Quanzhou Normal University*.
- [4] Liu Lizhi. Urban Image Construction and Lifestyle : A Case Study of Guangzhou's City Image and Guang-style Living. *Modern Urban Studies*, 2023(03).
- [5] Shi, Q. (2024). The Multimodal Discourse Analysis of Asian Games Videos from the Perspective of Visual Grammar. *Advances in Social Development and Education*.
- [6] Vojnović, D. V.(2020). Multimodal discourse analysis of tourism websites–The promotion of cultural values through the language of tourism to mark the 90th anniversary of the English.
- [7] Wang Yanqi. Analysis of Communication Effects and Influencing Factors of Urban Image: A Case Study of Chongqing's City Image Videos on YouTube. *Media*, 2024(23).
- [8] Xiao Jun. Multimodal Discourse Analysis: Theoretical Model and Its Methodological Significance for New Media Cross-Cultural Communication Research. *Journal of Wuhan University (Humanities and Social Sciences Edition)*.2017(04)
- [9] Xiao Jun and Zhang Chi. A Study on Symbolic Narratives in Cross-Cultural Communication of Short Videos. *Journalism and Writing*. 2024(03)
- [10] Yan Yan. Problems and Countermeasures of Urban Image Communication in the Era of Short Videos: A Case Study of Changsha's Urban Image Communication. *China Journal of Radio and Television*, 2023(04).
- [11] Yanwei, J., & Hassan, H.[1] Yanwei, J., & Hassan, H. (2023). A Multimodal Discourse Analysis of Tourism Promotional Video Beijing. *International Journal of Academic Research in Linguistics*.
- [12] Zahra, M. A. L. M. (2021). The Representation of Syrian Refugees from the Perspective of the Western Communities on Facebook: Corpus-based Multimodal Cognitive Critical Discourse.