

Translation as Image Crafting: Strategic Multimodal Highlighting in the Representation of Hangzhou's Image

Yuanxin Yang*, Xi'an Jiaotong-Liverpool University, University of Liverpool

Hui Wang**, Xi'an Jiaotong-Liverpool University

ABSTRACT

This article investigates how Hangzhou's bilingual city promotional videos function as soft power instruments through strategic multimodal highlighting. Informed by Kress and van Leeuwen's (2021) visual grammar, Krifka's (2008) information focus notion, and van Leeuwen's (1999) auditory semiotics, it proposes a multimodal highlighting framework to examine how visual, verbal, and auditory elements are selectively deployed and highlighted in the Chinese source and translated English videos to construct ideologically nuanced city images. The analysis reveals that the Chinese video favours rhetorical density and implicit evaluative phrasing, foregrounding domestic unity and resilience, while the English video prefers to employ explicit focus markers and pitch-based prominence to project global appeal and modernity. The shift from implicit resonance to explicit legibility underpins the ideological mediation shaped by audience orientation and geopolitical intent. By contextualising these multimodal highlighting strategies within China's state-led communication system, this article demonstrates that multilingual city branding operates as a dual-track ideological instrument, deeply integrated into broader practices of national soft power.

KEYWORDS

Multimodal highlighting strategies, city branding, soft power, translation and ideology, bilingual promotional videos.

1. Introduction

In today's global competition for investment, talent, and influence, cities increasingly shape their images on social media by leveraging symbolic resources to enhance global appeal (Dinnie, 2011; Wang & Feng, 2023). On the eve of the 19th Asian Games 2023, Hangzhou, a Chinese city with a 5,000-year history and now renowned as a leading "AI hub" (Fu, 2025), released bilingual promotional videos titled *What is Hangzhou?* on video-sharing platforms like YouTube, Bilibili and RedNote, promoting its identity as both a historically resonant cultural city and a modern hub of technological innovation. With over 186 million views (Hangzhou Municipal Bureau of Culture, Radio, TV, and Tourism, 2025), the videos integrate expertise from history, archaeology, urban planning and conservation to enrich Hangzhou's narrative of historical continuity and urban transformation. Unlike typical official city promotional videos that often prioritise visually aesthetic landscapes, Hangzhou's videos adopt a more "reflective and deconstructive" approach (Zhang, 2023), centred on the core

* ORCID 0009-0003-1952-8278, e-mail: Yuanxin.Yang@liverpool.ac.uk

** ORCID 0000-0002-3226-8728, e-mail (corresponding author): Hui.Wang@xjtlu.edu.cn



creative concept of “5000 Years, One City: Hangzhou”. Beginning with the fundamental question “What is Hangzhou?”, the narrative guides the audience through a journey blending history, geography, and technological innovation. It is interesting to note that the videos strategically employ linguistic, visual, and auditory resources through editing choices to foreground key aspects of Hangzhou’s identity and emotional appeal, shaping a multimodal narrative of its evolution towards the ‘Heaven on Earth’ and the ‘AI hub.’ This study examines how such multimodal resources are selectively shuffled and reshuffled to emphasise particular information, contributing to the construction and communication of Hangzhou’s distinctive city image.

Information in promotional videos is carefully selected to shape public perception of urban identity persuasively (Lang, 2024; Yao & Zhuo, 2018). Capturing attention is crucial for message processing and effective promotion (Wilson et al., 2015). However, how multimodal resources coordinate to direct attention toward key information remains underexplored. Linguistic studies on information structure focus mainly on syntactic and textual mechanisms of information emphasis (e.g., Krifka, 2008; Xu, 2004; Zubizarreta, 2019). Although semioticians have discussed information value and visual salience in still images (e.g., Kress & van Leeuwen, 2021; Machin & van Leeuwen, 2016), and marketing studies emphasise attention as central to persuasion (Mills, 2000; Simons, 1986), video-based multimodal attention-getting, or highlighting, remains insufficiently addressed. Limited research examines how such strategies are specifically tailored in city promotional videos to foreground distinctive features and establish a lasting, favourable impression. Furthermore, while consistently encoded grammatically, the forms of manifesting information focus vary across languages (Xu, 2004). Another layer of complexity emerges in bilingual or translated versions, where translation can reshape focal information and recast city images for audiences with diverse ideological backgrounds (Li & Pan, 2021). Yet, how multimodal focal content is reframed in the translation process to reshape a city’s image and align it with audience expectations remains underexplored.

Building on these discussions, this study adopts a multimodal critical discourse approach, which extends critical discourse analysis to the interplay of multiple semiotic modes and the ideological meanings they jointly construct (Machin, 2013), to examine the construction of Hangzhou’s city image in its Chinese promotional videos and the translated English version, with a focus on multimodal strategies of emphasis. The research questions include: 1) How are various aspects of Hangzhou’s image emphasised in the Chinese promotional video? 2) How are these emphases reshuffled and re-highlighted in the target English video? 3) What are the underlying factors behind the changes? By addressing these questions, this study investigates multimodal highlighting strategies in relation to their underlying sociopolitical motivations and communicative intentions.

The following sections first outline the theoretical framework, combining a micro-level analysis of multimodal highlighting strategies with a macro-level exploration of their underlying motivations. This is followed by an in-depth analysis of emphasising strategies adopted in Hangzhou's promotional videos. Finally, the findings are discussed in relation to the reshaping of Hangzhou's narrative and its implications for soft power projection.

2. Navigating cross-cultural communication: an integrated theoretical framework

Image is an indispensable part of national soft power and strategic capital (Feng, 2023; Keohane & Nye, 1998). China places considerable emphasis on global image-building to foster a more acceptable, empathetic, and neutral perception among international publics (D'Hooghe, 2015; Feng, 2023). Such promotional narratives go beyond introducing landmarks or economic power, but leverage on guiding attention to intended information for effective persuasion (Wilson et al., 2015). Their effectiveness hinges on how information is selected, framed, and conveyed across modes, involving foregrounding certain information while downplaying or omitting other information. These choices are embedded in both video production and multimodal expressions, aligning with the producer's intended value and projected image.

Given the growing use of audiovisual media in city branding and global communication (Feng, 2023), it is helpful to develop a systematic framework for analysing both multimodal mechanisms of information focusing and the broader sociocultural motivations underlying representational choices. To this end, this study proposes a dual-layered framework for exploring how multimodal highlighting strategies in bilingual promotional videos shape city images. At the micro level, it draws on Kress and van Leeuwen's (2021) visual grammar, Krifka's (2008) information focus, and van Leeuwen's (1999) audio grammar to analyse how visual, linguistic, and auditory resources are orchestrated to foreground specific elements. At the macro level, Nye's (1990) soft power theory provides an explanatory lens for interpreting the strategic intentions behind these representational choices. This integrated approach bridges multimodal discourse analysis with sociopolitical interpretation, offering both analytical precision and theoretical depth.

2.1. Micro level: a multimodal analytical framework for analysing highlighting strategies

Translation, viewed as a re-contextualisation process (Baker, 2006), involves not only linguistic conversion but also recalibration of meaning across multiple semiotic modes.

It functions as a soft power resource (Roig-Sanz et al., 2025), constituting a purposeful social practice engaging various stakeholders beyond translators alone (Wang & Zhu, 2009). Translating promotional videos entails foregrounding specific cultural and symbolic elements while downplaying or omitting others, shaping audience perception and enhancing persuasive impact. Such selective highlighting is crucial, as cities compete for global attention by emphasising distinct features (Alegro & Turnšek, 2021). Translation thus becomes “an offer of information”, tailored to align with the cultural and informational expectations of target audiences (Nord, 2018, p. 33). To examine how information selection affects city image highlighting at the presentation stage, this study develops a multimodal analytical framework (Figure 1). It integrates visual grammar, the notion of information focus, and audio grammar to identify and interpret highlighting strategies.

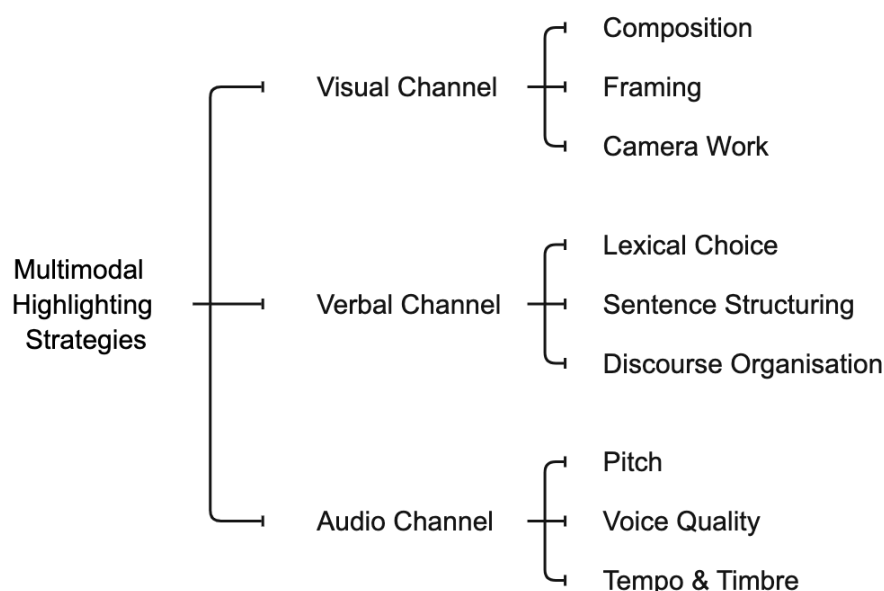


Figure 1 A multimodal analytical framework for analysing highlighting strategies

Core values are typically promoted visually in multimodal discourse, as audiences can rapidly process complex visual scenes, extract key elements, and retain them in memory over extended periods (Celikkale et al., 2013; Francesconi, 2011). To better understand how promotional videos visually guide audience attention, this study draws on Kress and van Leeuwen’s (2021) visual grammar, which adapts Halliday’s (1985) systemic functional linguistics to image analysis through the ideational, interpersonal, and textual metafunctions. Among these three metafunctions, the textual metafunction is the most directly relevant to the present study, because it concerns how visual elements are arranged into a coherent composition and how certain elements are foregrounded as focal. Such compositional meaning is realised mainly through the systems of information value, framing, and salience (Kress & van Leeuwen, 2021).

Information value concerns the placement of elements within the image, framing regulates the degree of connection or separation between them, and salience directs viewer attention through size, foregrounding and backgrounding, colour, and tonal contrast (Kress & van Leeuwen, 2021). Considering the dynamic nature of videos, this study also considers dynamic visual resources such as shot distance, camera angles, and movement as attention-getting techniques. In the interpersonal metafunction, different shot lengths and angles, as well as shifts between shot lengths, such as camera zooming, not only indicate social distance but also direct visual attention (Dai, 2015; Verstraten, 2009).

While core values are primarily conveyed visually, they are confirmed by accompanying verbal elements that direct attention and shape interpretation (Francesconi, 2011). This study adopts Krifka's (2008) information focus notion to investigate how emphasis is manifested linguistically, as focus plays a key role in orienting attention (Bornkessel-Schlesewsky & Schumacher, 2019). Drawing on prior studies of focus marking/construction in written and spoken languages (e.g., Chen et al., 2019; Krifka, 2008; Xu, 2004), this study categorised Chinese and English focus-marking strategies into lexical, sentence, and discourse levels (Figure 1). At the lexical level, focus-sensitive particles, such as *only*, *also*, and *even* in English, and *jiu*, *cai*, *zhi*, and *dou* in Chinese, can affect a sentence's truth condition (Chen et al., 2019; Krifka, 2008). For example, the exclusive particle *only* emphasises uniqueness (Krifka, 2008), while *jiu* serves as a restrictive focus particle in Chinese (Chen et al., 2019; Lu, 2018; Zhang & Zhang, 1998). Promotional boosts refer to words with positive or superlative connotations, such as *excellent* and *most*, which enhance the qualities of the promoted entity without grammatical devices (Torresi, 2021). At the sentence level, cleft constructions, including *it*-clefts and *wh*-clefts in English and *shi...de* clefts in Chinese, often "signal an exhaustive interpretation that in-situ focus lacks" (Krifka, 2008, p. 18). Word order also serves to direct focal attention in information structure (Tomlin, 1995; Jasinskaja, 2019). Syntactic patterns, such as *lian...ye/dou...* and *you* structures in Chinese, and *there be* structures in English, are identified as effective focus-marking or highlighting strategies (Chen et al., 2019; Lu, 2018; Wen, 2012; Zhang & Jiang, 2005). At the text level, rhetorical emphasis, such as parallelism, repetition, contrast, and personification, provides overt, easily recognisable forms of focus (Lu, 2018).

Sound is vital in shaping image and emotion in multimodal texts, acting as a narrative force alongside visuals (Verstraten, 2009). Informed by van Leeuwen's (1999) audio grammar, this study investigates how key information is reinforced acoustically, with particular attention to pitch and voice quality in voiceover, as well as to tempo and timbre in music. These resources shape audience perception and emotional engagement with video content (Machin & van Leeuwen, 2016; van Leeuwen, 1999). Pitch, referring to sound height (van Leeuwen, 1999), marks the most prominent word,

or nuclear pitch accent, through rhythmic contrasts, thus signalling information focus (Zubizarreta, 2019). Voice quality, involving features such as tension, roughness, breathiness, and loudness, conveys rich connotative meanings, facilitating shaping impressions, evoking emotional responses, and influencing perceptions of the speaker's credibility (van Leeuwen, 1999). Additionally, tempo contributes to highlighting not by pointing to a specific item but by creating rhythmic contrast, especially the contrast of the emotional tone of the scene: a faster tempo evokes excitement or urgency, thereby accenting climactic moments, while a slower pace induces calm or melancholy, foregrounding contemplative or relaxing sequences (Machin & van Leeuwen, 2016; van Leeuwen, 1999). Timbre, referring to the tonal colouring or quality of a sound, also plays a highlighting role through affective differentiation. A marked shift or modulation in timbre can lend prominence to a scene or signal a shift in narrative importance without altering lexical content (van Leeuwen, 1999).

By integrating verbal, visual, and audio dimensions, this multi-layered framework provides a range of signals for analysing how multimodal resources underscore key information. It enables detailed analysis of how promotional videos use multimodal highlighting strategies to engage audiences and construct emotionally resonant city images, promoting distinctive urban identities and enhancing persuasive power.

2.2. Macro level: soft power and cross-cultural communication

As Fowler (1991) argues, selection and foregrounding are inherently ideological, reflecting the selector's intentions and interpretive stance. In promotional videos, these multimodal highlighting strategies contribute to shaping global perceptions of a city, and by extension, a nation. This introduces Nye's (2004) theory of soft power, which offers a valuable perspective for understanding how cities enhance their domestic and global appeal and cultural influence through carefully organised bilingual media representations. From this perspective, multimodal highlighting becomes a tool for managing images and projecting soft power to craft a favourable global urban identity.

As a central element of soft power, image communicates perceptions of a nation's political system, economic development, cultural values, and societal characteristics (Feng, 2023). It reflects a country's intention to draw international attention to its achievements and foster belief in its national qualities (Jordan, 2014). This aligns with Rawnsley's (2012) conception of soft power as 'attractive' power, stressing the ideals and principles a nation upholds and seeks to promote globally. Originally proposed by Nye (1990, p. 167) as "getting others to want the outcomes that you want", soft power emphasises persuasion through attraction rather than coercion (Nye, 1990; Roig-Sanz et al., 2025). Its ultimate goal is to shape global perceptions by encouraging

acceptance of particular norms, values, lifestyles, and social systems through moral and emotional appeals (Meng, 2020).

Deploying soft power necessitates cultural sensitivity and engagement with others' values and motivations (Roig-Sanz et al., 2025). In city branding, tailored storytelling helps make the intended image accessible across cultural borders. As a soft power resource, translation is “a non-neutral activity that unfolds in specific social and political contexts” (ibid., p. 1). Translators employ diverse strategies to renegotiate narratives, aiming to “strengthen or undermine particular aspects of the narratives they mediate, explicitly or implicitly” to “produce a politically charged narrative in the target context” (Baker, 2006, p. 105). These narrative choices rest on the selective organisation of information, that is, deciding what to highlight, background, and how to direct attention. Such emphasis is intentional, foregrounding elements that both reflect the desired city image and resonate with the target audiences' values. In this light, linguistic, visual, and acoustic highlighting choices form multimodal tools essential for projecting soft power. They determine which aspects of the city stand out, how meaning is constructed, and how perceptions and emotional responses are guided. This reinforces the need to analyse how multimodal emphasis is orchestrated to project soft power in globally oriented promotional videos.

As soft power practices, city promotional videos selectively present a city's culture, heritage, achievements, and people to tell tailored stories enhancing identity and appeal domestically and internationally. To elucidate the factors governing information highlighting in image projection and focal shifts during the translation process, this study synthesises Nye's (1990) soft power theory with a micro-level framework of multimodal highlighting strategies, linking detailed textual mechanisms with broader sociopolitical motivations. This dual perspective clarifies how image construction serves broader city branding goals and lays a theoretical foundation for the case study.

3. Case analysis

3.1. Data selection and coding

The twelve-minute promotional video for Hangzhou, titled *What is Hangzhou?*, opens with a central question: how Hangzhou transformed its destiny three times to become today's “人间天堂” ‘Heaven on Earth’. It then traces Hangzhou's three major transformations over 5000 years from ancient civilisation to modern metropolis, culminating in its reinforced image as Heaven on Earth, an international innovation hub, and living poetry. The video was chosen because it departs from conventional scenic-collage official promotion, offering a hybrid, narrative-driven structure that makes multimodal highlighting strategies particularly salient. Moreover, its explicit bilingual

dubbing allows for systematic comparison of how emphasis is deployed across language versions. The dataset comprises two dubbed versions of the video with the same visuals: a Mandarin version with Chinese and English subtitles (TT1), narrated by a young, light female voice, and an English version with English subtitles (TT2) only, narrated by a mature and full female voice (Hangzhou Municipal Bureau of Culture, Radio, TV and Tourism, 2023a, 2023b). For detailed comparison, four excerpts were chosen for analysis: the opening remarks, the introduction of Hangzhou's three transformations, the birth of Liangzhu civilisation, and Hangzhou's modernisation and globalisation (see Table 1). They condense the video's four main image dimensions: Hangzhou as 'Heaven on Earth', a city shaped by resilient transformation, a site of deep civilisational heritage, and a modern city with global economic reach. Together, these excerpts amount to 51 seconds, 29 shot-based visual units, and 19 lines.

Excerpt	Timestamp	Length (s)	Theme	No. of Shots
1	00:00-00:07	7	The opening remarks	5
2	00:34-00:50	16	Introduction to Hangzhou's transformations	12
3	03:01-03:10	9	The birth of Liangzhu civilisation	2
4	10:02-10:21	19	The city's modernisation and globalisation	10

Table 1 Overview of the four selected excerpts for detailed analysis

The four excerpts were coded deductively in NVivo 15¹ using the mode-specific highlighting strategies discussed in Section 2.1. Visual strategies were coded in shot-based units directly on the video track, with a new unit identified whenever a cut or another clear visual transition introduces a different focal object or scene. Continuous camera movement within the same shot, such as a pan or zoom over the same object, was treated as a single unit as long as the main visual focus remains unchanged. Subtitles were then transcribed in NVivo and verbal and audio strategies were coded in the transcript, allowing verbal and prosodic features to be identified at the point of occurrence. Additionally, Praat² was used to inspect pitch movement and stress placement in selected excerpts. The separately coded verbal and audio features were then aligned analytically with the temporally corresponding visual unit(s) in the video. To enhance coding reliability, both authors independently coded a subset of the data using a predefined codebook. The coding results were then compared, and discrepancies were resolved by referring back to the operational definitions and the original data. The codebook was refined accordingly, and the first author then coded the remaining data. Ambiguous cases were further discussed and resolved with the second author where necessary.

3.2. Image framing in *What is Hangzhou?*

3.2.1. Multimodal highlighting overview across four excerpts

Following the theoretical framework developed in Section 2.1, this section provides a comparative overview of the highlighting strategies identified in the four selected excerpts from the Chinese and the English versions (see Table 2).

Modes	Highlighting strategies	Chinese-dubbed video	English-dubbed video
Visual channel	Saliency		16
	Camera movement		8
	Information value		6
Verbal channel	Focus-sensitive operators	3	6
	Lexical boosts	1	3
	Emphatic words	3	0
	Rhetorical emphasis	8	1
Audio channel	Stress	17	14
	Prolonged duration	12	9
	High pitch	1	12

Table 2 Comparison of highlighting strategies in the two versions of the selected excerpts

Visually, three key strategies are adopted to direct attention in both versions: (a) saliency (e.g., foregrounding and size contrast); (b) dynamic camera movement (e.g., zoom and fast motion); and (c) centre-margin composition (e.g., central placement of focal objects or scenes). They highlight key city symbols, historical references, and scenes of urban modernity, providing a shared visual baseline for shaping Hangzhou's image.

The major divergence emerges in the verbal and audio channels. Roughly, the Chinese version stresses more verbal highlighting, while the English version relies more on auditory emphasis. Verbally, the Chinese version employs 15 distinct strategies, with *rhetorical emphasis* being the most frequent, appearing eight times. Conversely, the English version utilises ten strategies, including six instances of *focus-sensitive operators* and a total absence of emphatic words. This distribution suggests that the Chinese version tends to foreground meaning through rhetorical build-up and evaluative phrasing, whereas the English version relies more heavily on increasing informational explicitness of key messages.

Aurally, three strategies are identified across both versions, though they are distributed more evenly in the English video. While *stress* serves as the primary mechanism in both versions, a significant disparity exists regarding *high pitch*: it appears only once in the Chinese version, compared to twelve occurrences in the English video. Combining a light, young, and more flowing Mandarin voiceover, as opposed to a full, mature and assertive English voiceover, this distribution indicates that the Chinese version marks focal information with a smoother, more compact pitch pattern, while the English version relies much more on pitch-based prominence.

Overall, the two versions present distinct verbal and aural highlighting strategies. A detailed multimodal analysis in the following sections shows how these highlighting strategies operate in the videos to strengthen Hangzhou's identity.

3.2.2. Hangzhou: poetic framing vs. landmark identification

Excerpt 1 lasts seven seconds and contains five shot-based visual units, organised around four spoken lines (Table 3). Visually, the excerpt follows the same compositional patterns and colour schemes across the four lines. It opens with a full-screen view of West Lake, Hangzhou's iconic landscape. As shown in the 2nd visual, the shot is reframed through an inward-moving circular composition. This centred framing and the light grey background make the landmark the most salient visual element here. The shots across Lines 2-4 use the same central-circle composition and light-dark colour contrast to draw attention to the central photographs of landmarks. Meanwhile, bold, dark-coloured, large calligraphic characters (“美” ‘charm’, “潮” ‘river tide’, “禅意” ‘Buddhist mood’, and “繁华” ‘prosperity’ in Table 3) fall from the top of the scenes, which are partially obscured but still legible, shifting attention to these evaluative attributes. The contrast between realistic foreground photography and the abstract ink-wash background accentuates the juxtaposition of Hangzhou's historical charm and modern vitality. Therefore, the visuals couple each place with an evaluative attribute, directing attention towards the qualities through which they are to be perceived.

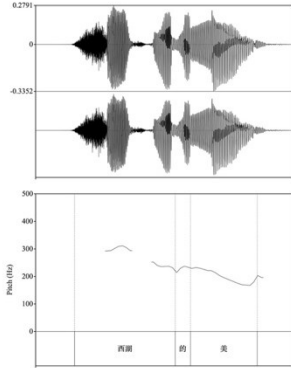
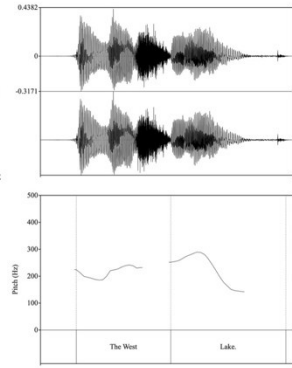
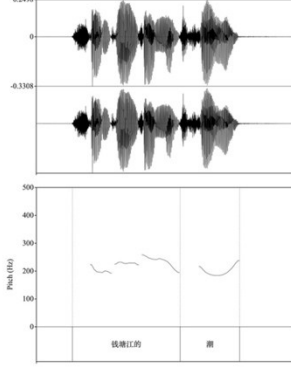
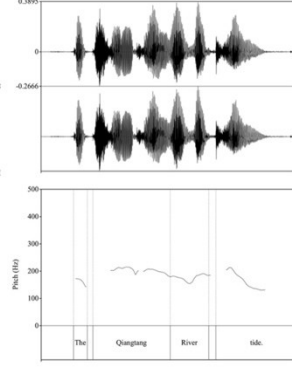
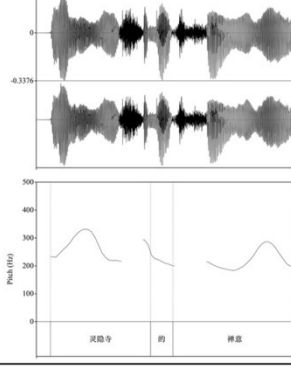
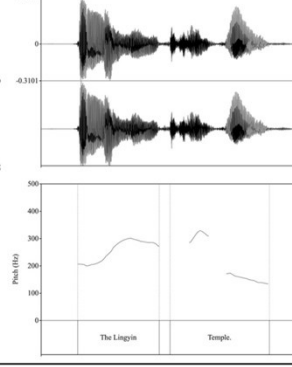
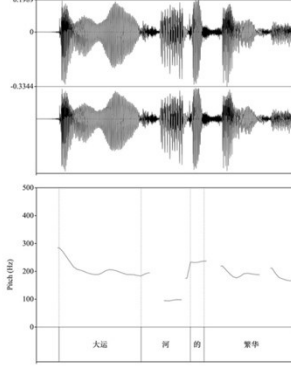
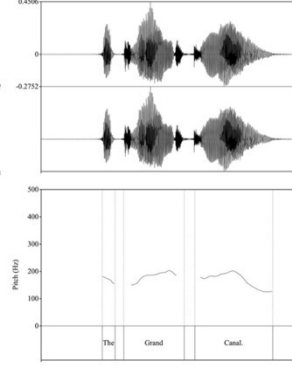
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
1		ST: 西湖的美 TT1: The West Lake's charm. TT2: The West Lake.		
2		ST: 钱塘江的潮 TT1: The Qiantang River's tidal bore. TT2: The Qiantang River tide.		
3		ST: 灵隐寺的禅意 TT1: The Lingyin Temple's serenity. TT2: The Lingyin Temple.		
4		ST: 大运河的繁华 TT1: The Grand Canal's prosperity. TT2: The Grand Canal.		

Table 3 Excerpt 1: Multimodal alignment (Note: ST=Chinese source text; TT1=English subtitles in the Mandarin version; TT2=English dubbing in the English version. The “CH Soundtrack” and “EN Soundtrack” columns display spectrograms illustrating the acoustic features of the Chinese and English dubbed utterances respectively.)

This evaluative orientation is reinforced in the Chinese verbal. Across four parallel lines, the ST adopts the same ‘attributive + noun’ structure to build the language’s momentum and arouse the audience’s affective response (Lu, 2018). The headwords (“美” ‘charm’, “潮” ‘river tide’, “禅意” ‘Buddhist mood’, and “繁华” ‘prosperity’) are focal here, because sentence-final positions in Chinese typically bear information focus (Xu, 2004). The verbal emphasis, therefore, falls on the distinctive qualities associated with the landmarks. This pattern closely corresponds to the visual design, where the landmarks occupy the compositional centre while the evaluative characters remain visible. These evaluative headwords also become aurally prominent with longer duration and noticeable pitch movement, such as the falling pitch contour on “美” ‘charm’ in Line 1 (Table 3), which draws attention to the descriptive endpoint instead of West Lake itself. These audio cues work together with the clause-final placement of the evaluative headwords and the visual salience of the landmarks, producing a tightly aligned multimodal focus on Hangzhou’s beauty, vitality, serenity, and prosperity.

TT1 preserves the focal structure of the ST by retaining the evaluative endpoints, so that the landmarks continue to be presented alongside the qualities that define their significance. TT2, by contrast, systematically shifts the excerpt’s focal structure. The evaluative headwords in ST are omitted in all four lines, and only the geographical referent is retained, such as “West Lake”. This change removes the information focus in the ST and TT1, and TT2’s focus becomes the landmarks’ identifiable place names. Aural stress and pitch prominence also fall on the clause-final nouns “Lake”, “River”, “Temple”, and “Canal”, as illustrated in Table 3, further reinforcing these place names and making them the audio foci. As a result, Hangzhou is reshaped from a culturally resonant and affectively textured city into a more identifiable, destination-oriented urban image.

3.2.3. Three pivotal transformations: resilient vs. prosperous

Excerpt 2 lasts 16 seconds and contains twelve shot-based visual units, organised around five spoken lines. Table 4 displays eight representative shots; four units that use similar highlighting strategies are omitted due to space constraints. The excerpt frames Hangzhou’s development as a movement from hardship to prosperity. Visually, the excerpt employs rapid forward motion through traditional gateways, buildings, and gardens (Table 4, Line 1), suggesting a passage through historical time. In Lines 2-5, the visuals shift to more stable aerial shots of city sceneries, foregrounding present

prosperity alongside historical grandeur. The transitions between shots across Lines 2-5 are rendered through an ink-wash-like diffusion effect, resembling ink slowly spreading on Xuan paper, highlighting the video's broader projection of Hangzhou as a misty, poetic Jiangnan city. The visual progression here supports a narrative of endurance, transformation, and arrival.

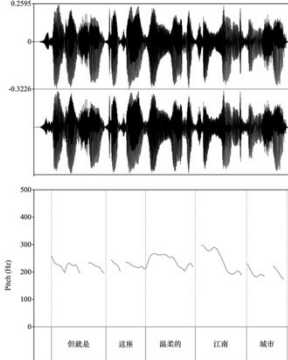
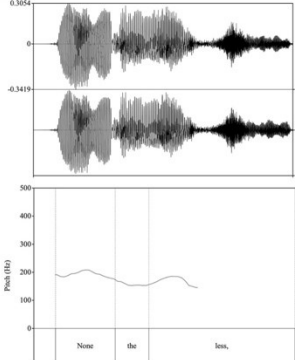
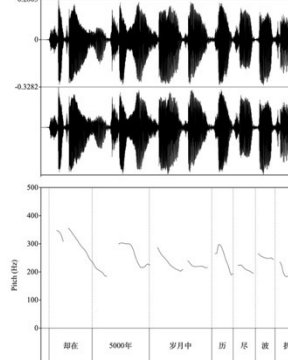
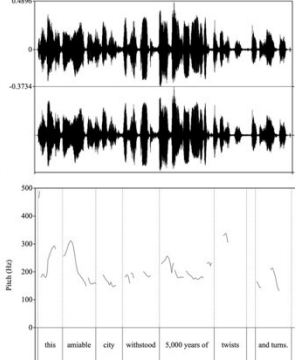
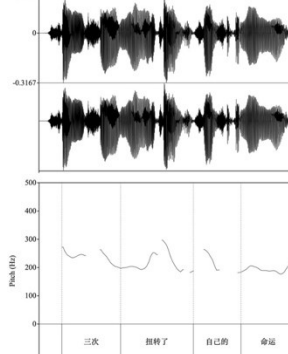
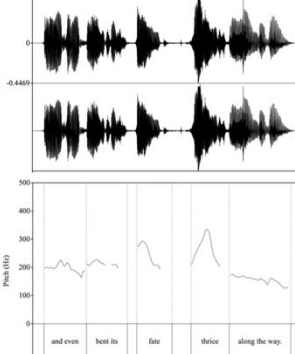
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
1		ST: 但就是这座温柔的江南城市 TT1: Nonetheless, this amiable city in the Jiangnan region TT2: Nonetheless,		
2		ST: 却在 5000 年岁月中历尽波折 TT1: withstood 5,000 years of twists and turns TT2: this amiable city withstood 5,000 years of twists and turns		
3		ST: 三次扭转了自己的命运 TT1: and even managed to bend its fate thrice along the way. TT2: and even bent its fate thrice along the way.		

Table 4 Excerpt 2: Multimodal alignment

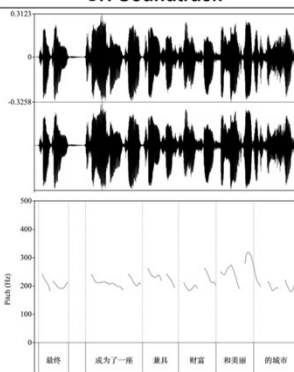
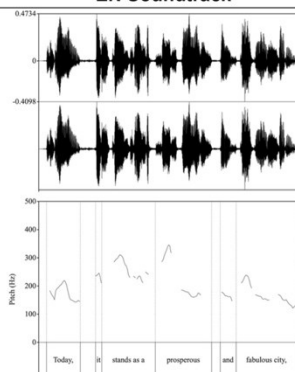
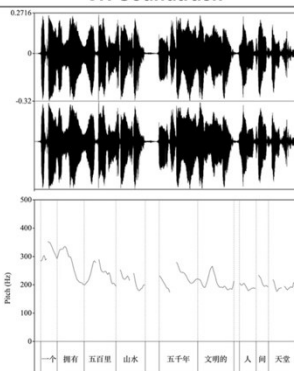
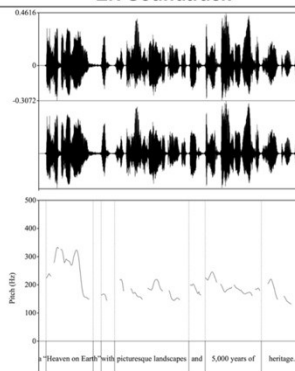
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
4		ST: 最终 成为了一座兼具财富和美丽的城市 TT1: It eventually became a prosperous and fabulous city, TT2: Today, it stands as a prosperous and fabulous city,		
5		ST: 一个拥有五百里山水 五千年文明的人间天堂 TT1: a 'Heaven on Earth' with picturesque landscapes and 5,000 years of heritage. TT2: a 'Heaven on Earth' with picturesque landscapes and 5,000 years of heritage.		

Table 4 (Continued)

The visual framing is reinforced in the ST, particularly through rhetorical devices. In Line 1 (Table 4), Hangzhou is introduced as “*温柔的江南城市*” ‘amiable Jiangnan city’, a personified formulation that foregrounds its gentle regional identity. The contrastive opening “*但就是*” ‘nonetheless’ marks what follows as noteworthy. In Lines 2 and 3, the ST shifts to adversity and agency through “*历尽波折*” ‘withstood twists and turns’ and “*三次扭转...命运*” ‘bent its fate three times’. These expressions personify the city and present its history as an active struggle against hardship. Acoustically, “*5000 年*” ‘5000 years’, “*历尽波折*” ‘withstood twists and turns’, and “*三次*” ‘three times’ are pronounced with stress and prolonged duration, and noticeable pitch movement, making them aurally prominent. In Lines 4 and 5, the focus moves to the outcome of the evolution. The resultative formulation “*最终成为*” ‘eventually become’ directs attention to present achievement, while the numerical parallelism of “*五百里*” ‘500 li, 500 metres per li’ and “*五千年*” ‘5000 years’, which are also acoustically stressed, foregrounds Hangzhou’s scenic breadth and civilisational depth.

TT1 largely preserves the ST’s focal progression. It retains the contrastive opening “Nonetheless,” the evaluative description “this amiable city in the Jiangnan region”, the resilience-oriented phrasing “withstood...turns” and “managed to bend its fate” that foregrounds effort and hard-won achievement, and the resultative framing “eventually

became". It broadly reproduces the ST's processual narrative, presenting Hangzhou as a gentle yet resilient city whose prosperity results from endurance and transformation. TT2, however, recalibrates the excerpt's focal structure. It retains much of the core content but restructures the progression. Line 1 isolates "Nonetheless," as a standalone attention unit before the city is named in Line 2. Line 4 is also reorganised as "Today, it stands as", replacing the ST's resultative, past-oriented framing with a present-oriented one. This shift weakens the sense of historical process and foregrounds the current status instead. Compared with the smoother Mandarin delivery, the English dubbing uses a wider pitch range and more clearly segmented stress patterns, with marked prominence on present-oriented and evaluative items such as "Today," "prosperous", and "Heaven on Earth". Therefore, resilience is not removed from the English version but is subordinated to a stronger emphasis on present-day success and recognisable achievement.

3.2.4. The birth of Liangzhu civilisation: pioneering vs. heritage-rich

Excerpt 3 features a nine-second duration, containing two shot-based visual units and three spoken lines (Table 5), presenting Liangzhu culture as a civilisational origin and part of a wider global historical frame. Visually, in the first shot (Table 5, Lines 1 & 2), a central close-up of the yellow jade Cong, a jade annular object from the Liangzhu period, surrounded by other jade artefacts in the dark background and illuminated by a golden spotlight on a circular platform, gives the Cong high salience. In the second shot (Line 3), the video shifts to a stylised backdrop of the four ancient civilisations, with the Cong placed among them. Although all four civilisations are shown together, the Chinese section is rendered with denser lines, more intricate patterns, and a stronger glow, which subtly makes it the visual focus. The visuals thus frame Liangzhu as a civilisation positioned within a global historical order.

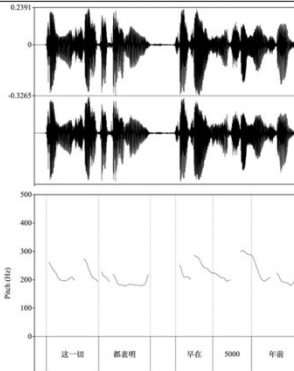
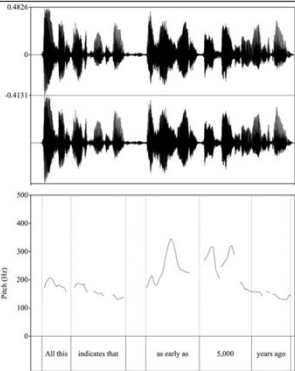
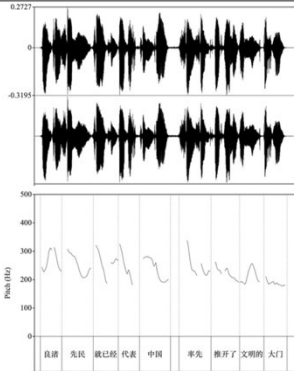
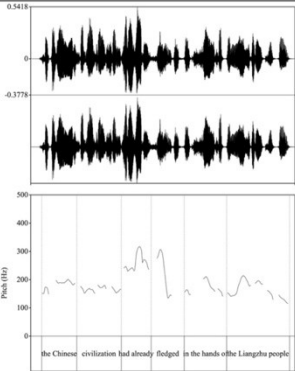
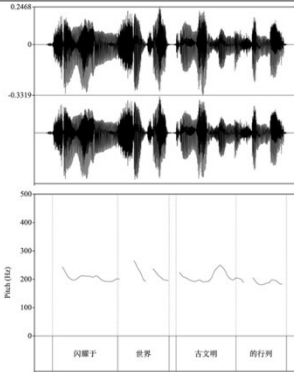
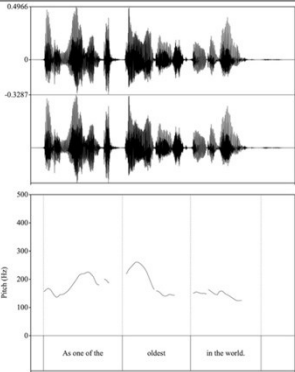
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
1		ST: 这一切都表明早在 5000 年前 TT1: All this evidence indicates that, as far back as 5,000 years ago, TT2: All this indicates that as early as 5,000 years ago		
2		ST: 良渚先民就已经代表中国 率先推开了文明的大门 TT1: the Chinese civilization was already budding in the hands of the Liangzhu people TT2: the Chinese civilization had already fledged in the hands of the Liangzhu people		
3		ST: 闪耀于世界古文明的行列 TT1: as one of the oldest in the world. TT2: as one of the oldest in the world.		

Table 5 Excerpt 3: Multimodal alignment

Echoing the visual emphasis, in the ST of Line 1, the focus-sensitive adverb “都” ‘all’ highlights the previously shown evidence as collectively proving Liangzhu’s antiquity, while “早在” ‘as early as’ foregrounds the following time expression “5000 年前” ‘5000 years ago’, of which “5000” is aurally stressed and prolonged, making the long historical timespan especially prominent. In Line 2, “就” ‘so far as’ and “已经” ‘already’ further mark precedence, while “率先” ‘pioneer’ is acoustically marked to emphasise that the Liangzhu people were the first to “推开...大门” ‘open the door of...’, giving Liangzhu an explicitly pioneering role in Chinese civilisation. In Line 3, the stressed verb “闪耀” ‘shine’ echoes the visually shining Chinese civilisation section, while the stressed “世界” ‘world’ extends the significance of Liangzhu culture to the global

context. Combining the comparative visual arrangement with these verbal and audio strategies, Liangzhu culture is positioned as pioneering, shining among the world's ancient civilisations.

TT1 maintains part of the ST's focal structure while simultaneously reconfiguring its emphasis. In Line 1, the use of "All this evidence" and "as far back as" retains the focus on antiquity and more explicitly articulates the evidential relationship between the previously mentioned evidence and the fact that Chinese civilisation had already begun to emerge in the Liangzhu period. In Line 2, however, the ST's agentive metaphor of "opening the door of civilisation" is replaced by "the Chinese civilization was already budding...". The predicate restructuring weakens the sense of Liangzhu as an active pioneer, shifting the attention towards Chinese civilisation's developmental status. Additionally, Line 3 is reduced to "one of the oldest...", which narrows the ST's emphasis to historical age. TT2 accentuates this shift in focus. Likewise, it preserves the antiquity marker "as early as". But "the Chinese civilization had already fledged" in Line 2 foregrounds a resultative and mature state by changing TT1's "budding" to "fledged", while the restructured predicate "had already" reinforces the idea that this state had been achieved prior to the relevant past reference point. Through predicate restructuring and morphosyntactic shift, TT2 further distances itself from the ST's original emphasis on pioneering agency. In the English dubbing, stress and pitch movement fall on "5000", "already fledged", and the superlative "oldest", reinforcing antiquity and established status rather than the act of opening civilisation. Similar to TT1, omitting "闪耀" 'shine' and "世界古文明" 'world's ancient civilisations' diminishes the ST's positive evaluation and shifts attention to age. As a result, the English dubbing shifts the emphasis from pioneering agency to antiquity and Hangzhou's established heritage.

3.2.5. Modernisation and digital economy: innovative vs. globalised

The 19-second Excerpt 4 consists of ten shot-based visual units and seven spoken lines (Table 6). It traces Hangzhou's development from traditional private enterprise and manufacturing to the modern digital economy. Visually, the excerpt opens with a top-bottom composition with a high-speed rail cutting across the frame, directing attention from foregrounded farmland to the skyline behind, indicating Hangzhou's transition from agriculture to urban modernity. The subsequent shots move through aerial views of Hangzhou's industrial clusters and fast-motion shots of the busy logistics industry, underscoring the city's urban prosperity.

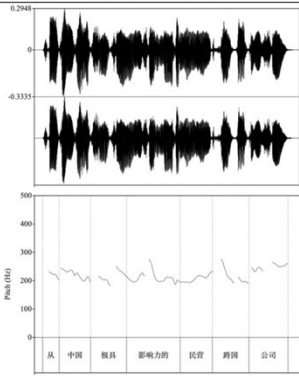
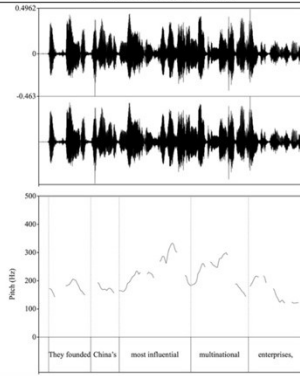
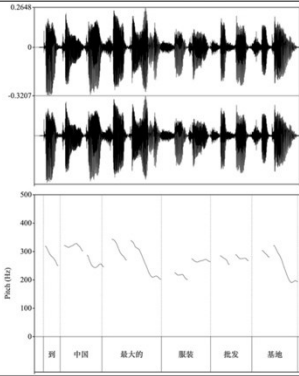
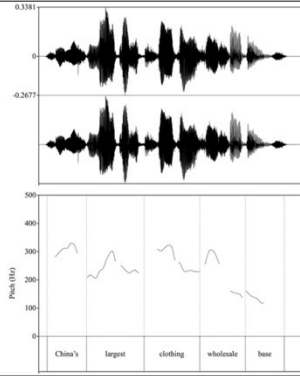
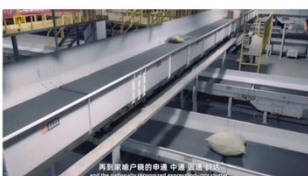
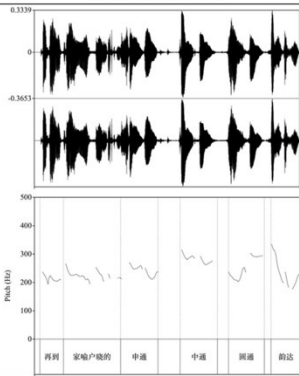
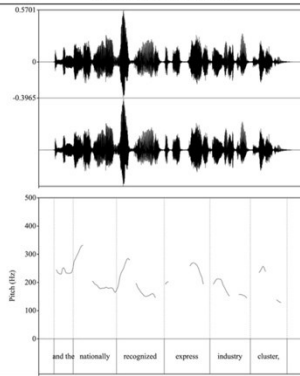
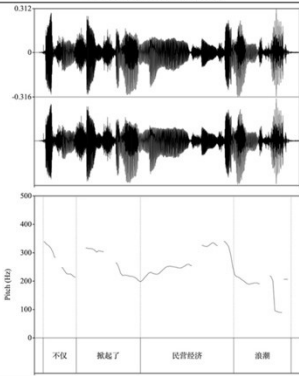
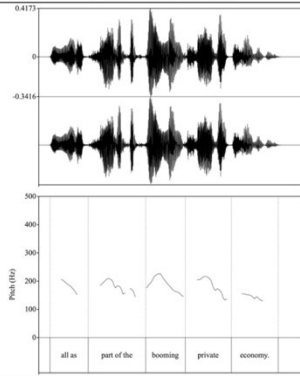
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
1		ST: 从中国极具影响力的民营跨国公司 TT1: They founded some of China's influential multinational enterprises, TT2: They founded China's most influential multinational enterprises,		
2		ST: 到中国最大的服装批发基地 TT1: China's largest clothing wholesale base, TT2: China's largest clothing wholesale base,		
3	 	ST: 再到家喻户晓的申通 中通 圆通 韵达 TT1: and the nationally recognized express industry cluster, TT2: and the nationally recognized express industry cluster,		
4	 	ST: 不仅掀起了民营经济浪潮 TT1: all as part of the booming private economy. TT2: all as part of the booming private economy.		

Table 6 Excerpt 4: Multimodal alignment

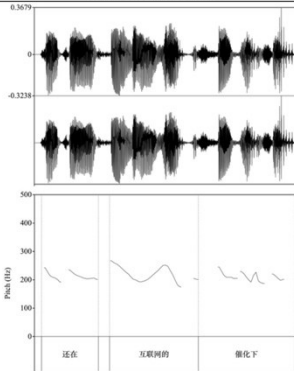
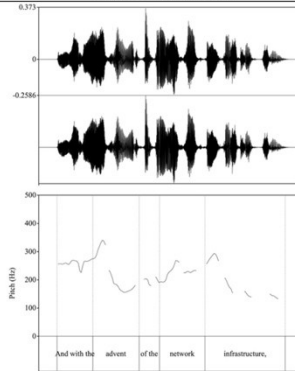
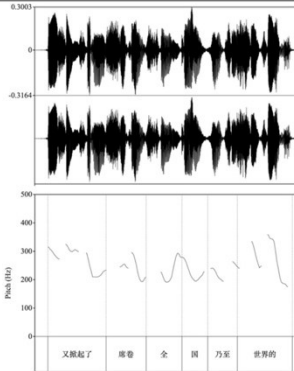
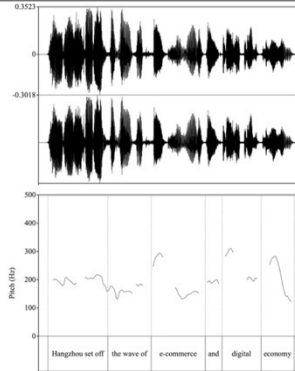
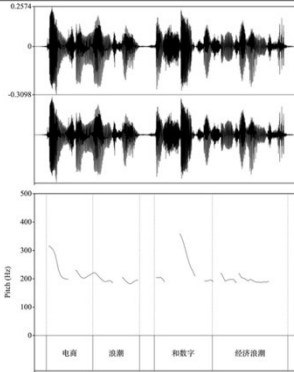
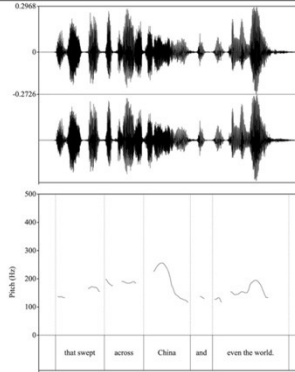
No.	Visual Image	Line	CH Soundtrack	EN Soundtrack
5		ST: 还在互联网的催化下 TT1: And with the advent of the network infrastructure, TT2: And with the advent of the network infrastructure,		
6	 	ST: 又掀起了席卷全国乃至世界的 TT1: Hangzhou set off the wave of e-commerce and the digital economy TT2: Hangzhou set off the wave of e-commerce and the digital economy		
7		ST: 电商浪潮和数字经济浪潮 TT1: that swept across China and the world. TT2: that swept across China and even the world.		

Table 6 (Continued)

In the ST, Lines 1-4 foreground Hangzhou's strength in private enterprise, manufacturing, and logistics. Promotional boosts such as “极具影响力” ‘extremely influential’ and “中国最大的” ‘China's largest’ place strong evaluative emphasis on influence and scale by superlatives. Hangzhou's urban development is further extended by the progressive structure “不仅...还...” ‘not only...but also...’ (Zhuang, 2015), which presents the digital economy as an expansion of earlier private-sector growth rather than a separate achievement. Acoustic prominence falls on phrases such as “极具影响力” ‘extremely influential’, “最大的” ‘the largest’, and “民营跨国公司” ‘privately owned multinational enterprises’, making influence, scale, and ownership aurally salient. In Lines 5-7, the ST shifts to the scope and scale of Hangzhou's digital

modernity through aurally stressed words “席卷” ‘sweeping’, “全国” ‘nationwide’, “世界” ‘world’, and the range-extending “乃至” ‘even to’. These choices frame Hangzhou not only as a successful local economy but as a driving force in nationwide and global digital transformation.

TT1 largely maintains the ST’s developmental sequence but softens its claims with a more measured voice. In Line 1, “some of” narrows the scope of the ST’s broader emphasis on highly influential private multinationals. The omission of “民营” ‘privately owned’ in TT1 also removes the ownership-based contrast with 国有 ‘state-owned’ that is salient for Chinese viewers. Likewise, although TT1 retains references to e-commerce and the digital economy that spreads “across China and the world”, it omits the progressive structure and breaks the lines into two separate sentences, thus reducing the ST’s progressive build-up from local private enterprise to global digital reach. Overall, TT1 presents Hangzhou’s economic importance and modernity, but in a more measured and less assertive way than the ST.

TT2 shifts the focal structure further towards global competitiveness. In Line 1, the superlative “most” strengthens the claim that Hangzhou is home to “China’s most influential multinational enterprises”. In Line 7, the insertion of the focus-sensitive operator “even” before “the world” heightens the degree of emphasis on global expansion. The English dubbing also gives prosodic prominence to scale- and reach-oriented items like “most influential”, “largest”, “multinational”, and “world.” Together, TT2 places greater emphasis on Hangzhou’s outward-facing success, recasting it less as a city rooted in the domestic rise of private enterprise and more as an internationally competitive centre of digital economy.

3.2.6. Summary of the four excerpts

Overall, the four excerpts demonstrate that the Chinese-dubbed version relies more heavily on verbal highlighting, especially *rhetorical emphasis*, as identified in Section 3.2.1. In the detailed analyses, the ST, alongside visual and audio representations, repeatedly attaches culturally resonant and affectively charged qualities and values to landmarks, historical references, and economic development through parallelism, personification, and emphatic expressions. As a result, Hangzhou’s qualities such as poetic charm, resilience, civilisational significance, and locally rooted dynamism are brought to the fore, constructing a culturally resonant and narratively textured image. Embedded in the Chinese video, TT1 remains relatively close to the ST’s focal structure. Although some rhetorical density is reduced, TT1 generally preserves the ST’s evaluative endpoints, temporal framing, and processual progression, allowing non-Chinese viewers to gain an impression of the city that is largely faithful to the source text.

By contrast, the English-dubbed version (TT2) relies more on *focus-sensitive operators* (six times), *lexical boosts* (three times), and, in particular, the pitch emphasis (twelve times). These strategies are more often applied to place names, historical status, superlative claims, and words signalling scale or global reach. In doing so, TT2 reorganises focal information by reducing culture-specific or rhetorically layered meanings and giving greater prominence to landmarks, established heritage, present achievement, and global competitiveness. As a result, TT2 reframes Hangzhou from a culturally resonant city into one that is more readily perceived as identifiable, accomplished, and globally competitive. Taken together, the Chinese video constructs a Hangzhou that is affectively rich and narratively dense, while the English video shifts emphasis towards tangible achievements and global standing.

4. Discussion

China has long endeavoured to present “a true, multi-dimensional and panoramic view of China” to a global audience (Xinhua News Agency, 2021). This objective accounts for the dual-language branding approach adopted by Chinese municipalities. As a quintessential hub of China’s digital economy (Fu, 2025), Hangzhou gained significant global resonance with its promotional film *What is Hangzhou?*, which successfully garnered millions of views worldwide. Co-produced by the Hangzhou municipal government and the Institute for Planets, a social media collective, the videos depart from the formulaic scenic-collage format often associated with official destination promotion that typically displays economic success, cultural flourishing, and social stability (Alegro & Turnšek, 2021; Feng, 2023). Instead, this collaborative model gives city branding a more hybrid and depoliticised appearance, blending official goals with civilian-style storytelling and affective appeal (Feng, 2023). Translation, in this process, operates as a process of multimodal (re)focalisation that (re)shapes what makes Hangzhou memorable, persuasive, and ideologically meaningful.

The Chinese-dubbed version foregrounds Hangzhou primarily through *rhetorical emphasis*, highlighting evaluative and culturally resonant qualities, values, and affective associations, and is acoustically supported by a narrower pitch range. This is because deploying specific rhetorical devices can effectively evoke profound, culturally specific connotations and a sense of pride in Chinese viewers, thereby emphasising ‘Chineseness’ (Guan & Forceville, 2020). By contrast, the English-dubbed version gives greater prominence to more explicit, recognisable focal attributes through *focus-sensitive operators*, *lexical boosts*, and pitch-based emphasis, as English writing prefers to use more explicit, direct assertions (Liao & Chen, 2009). The result is a shift in projected city images: from a culturally resonant, narratively dense portrayal to one that is more identifiable, outward-facing, and internationally legible. This difference is

tied to the different statuses of the ST and the two TTs. Embedded in the Chinese-dubbed video, TT1 allows non-Chinese viewers to access the source-oriented image logic while remaining subordinate to the Chinese audiovisual flow. TT2, by contrast, functions as the main verbal track of the English-dubbed version and is therefore more strongly shaped by the need to make Hangzhou accessible and persuasive for international viewers.

These cross-version differences can also be understood through the lens of high- and low-context cultures (Hall, 1976). In high-context Chinese discourse, connections, cultural references, and values are often left to be inferred by readers (Hu et al., 2024). Persuasive messages are conveyed through hints rather than being explicitly spelt out in Chinese discourse, expecting the audience to infer meanings and connotations (González et al., 2001; Matalene, 1985). As Guan and Forceville (2020) argue, the intended meaning in bilingual promotional materials, particularly the culturally specific undertones, is not equally accessible to all but rather depends on the audience's familiarity with the target culture, thereby making the Chinese version's rhetorical strategies effective for a Chinese viewership. On the contrary, low-context English communication tends to make logical relations, evaluative stance, and textual connections more explicit (Hu et al., 2024). From this perspective, TT2's reduction of culturally and rhetorically dense expressions can be seen as aligning with target cultural norms, which is crucial to better serve the target audience's understanding and viewing experience (Hogg et al., 2014). However, this alignment also entails omitting certain culturally specific affective nuances and rhetorical layering, thereby subtly reorienting ideological undertones without overtly altering factual content.

These highlighting differences are central to how Hangzhou's image is projected for different audiences. Although the Chinese government seeks to project Hangzhou as both modern and historic (Lang, 2024), the Chinese-dubbed version's multimodal highlighting does not simply balance the two but instead coheres into a predominantly linear account of Hangzhou's historical evolution and contemporary achievement. By foregrounding local people's overcoming adversity and changing their destiny, it reinforces a national development discourse of self-improvement and collective progress. This narrative aligns with China's audience-centred branding approach, which prioritises domestic recognition to stimulate national sentiment and reinforce political legitimacy (Pu, 2019). Together with a young, non-professionally trained female voice deliberately chosen to resonate with young, digital-native audiences (Institute for Planets, 2023), the Chinese version's emphasis on historical continuity and modern transformations helps bolster national pride and consolidate the Communist Party of China's legitimacy (Keane et al., 2021). Thus, the Chinese version mobilises cultural storytelling as an ideological tool, resonating with the soft-power strategy of cultivating loyalty and unity at home. What is worth noting, however, is that

this domestically oriented ideological work is not entirely sealed off from the embedded English subtitles (TT1). While TT1 remains subordinate to the Chinese audiovisual flow, it makes the same narrative of resilience and progress accessible to non-Chinese viewers, potentially extending the ideological appeal beyond the domestic audience.

By contrast, the English-dubbed version recalibrates Hangzhou's dual identity through a more modular, explicit form of image construction. Its stronger emphasis on informational explicitness shifts attention away from local struggle towards historical depth, modernity, and global integration. This reorientation reflects China's broader effort to move beyond earlier reticence in international communication and present itself through innovation, openness, and competitiveness (Narins & Agnew, 2020). By highlighting Hangzhou's global competitiveness and outward-facing profile, the English video contributes to China's image as "an innovative nation and a technologically advanced society" (Keane et al., 2021, p. 159), while also resonating with wider attempts to reshape the global image of Digital China and expand discursive and soft power in global media (Feng, 2023; Keane et al., 2021). Importantly, however, this recalibration does not simply delete cultural heritage. The English version repackages it in forms that are more internationally legible and more easily aligned with a non-confrontational yet globally appealing national image (Pu, 2019; Zhang, 2011). Thus, what changes is not whether Hangzhou is historic or modern, but which aspects of that dual identity are made most salient to international viewers. In this sense, the English version remains embedded in a broader soft-power framework, one also consistent with the retrospective affirmation of progress and the forward-looking vision of rejuvenation captured in the idea that "China's dream is an urban dream" (Taylor, 2015, p. 107).

In this light, the transition of multimodal highlighting functions as a form of ideological mediation. The Chinese-dubbed version foregrounds poetic charm, historical resilience, civilisational depth, and locally rooted transformation, presenting an image of China as culturally continuous, collectively resilient, and driven toward modern development. It promotes a China-oriented ideology of cultural confidence, national cohesion, and developmental legitimacy, persuading through resonance rather than coercion (D'Hooghe, 2015; Feng, 2023; Zhu, 2022). In the English-dubbed version, by contrast, ideology is mediated more subtly but no less decisively through recalibrated emphasis: poetic charm shifts to accessible cultural heritage, local struggle shifts to present achievement, and domestic industrial dynamism shifts to global innovation. What is promoted here is an outward-facing image of China as open, innovative, technologically advanced, and globally competitive, while still retaining recognisable civilisational heritage. These two ideological mediations are not simply parallel alternatives. Instead, this multilingual city branding operates as a dual-track ideological

instrument: one message for internal consumption, another for external persuasion, yet both ultimately serve the larger project of state-led soft power.

5. Conclusion

This study demonstrates how multimodal highlighting strategies serve as instruments for ideological mediation and soft power projection by emphasising particular features of Hangzhou in its bilingual city promotional videos. Specifically, the Chinese version foregrounds domestic unity and resilience through rhetorical density and implicit evaluative phrasing (e.g., parallelism, personification), while the English version projects global appeal and technological modernity via explicit focus markers and pitch-based prominence. This shift from implicit resonance to explicit legibility reflects strategic decisions shaped by intended audiences and broader ideological agendas. By revealing how translation operates as a multimodal (re)focalisation that (re)shapes what is made memorable, persuasive, and ideologically meaningful, this study shows that city image construction is a site where translation, discourse, and state-led soft power converge. Furthermore, the distinction between the China-oriented and international-oriented versions underscores the dual-track nature of contemporary Chinese soft power projection: one cultivating internal cohesion and cultural confidence, the other projecting outward competitiveness and innovation. In framing translation as a discursive practice embedded in power relations, this study offers critical insight into the ideological work performed by multimodal communication in city branding. It also contributes to the broader debate on how multilingual place branding operates as a strategic tool in state-led image management, with implications for other cities seeking to balance local cultural resonance and global legibility.

References

- Alegro, T., & Turnšek, M. (2021). Striving to be different but becoming the same: Creativity and destination brands' promotional videos. *Sustainability*, 13(1), 139. <https://doi.org/dx.doi.org/10.3390/su13010139>
- Baker, M. (2006). *Translation and conflict: A narrative account*. Routledge.
- Bornkessel-Schlesewsky, I., & Schumacher, P. B. (2019). Towards a neurobiology of information structure. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 581–598). Oxford University Press. <https://doi-org.liverpool.idm.oclc.org/10.1093/oxfordhb/9780199642670.013.22>
- Celikkale, B., Erdem, A., & Erdem, E. (2013). Visual attention-driven spatial pooling for image memorability. *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 976–983. <https://doi.org/10.1109/CVPRW.2013.142>
- Chen, Y., Po-lun Lee, P., & Pan, H. (2019). Topic and focus marking in Chinese. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 733–752). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642670.013.34>



- Dai, J. (2015). *电影批评 (Film criticism)* (2nd ed.). Peking University Press.
- D'Hooghe, I. (2015). *China's public diplomacy*. Brill Nijhoff.
- Dinnie, K. (2011). Introduction to the theory of city branding. In K. Dinnie (Ed.), *City Branding: Theory and Cases* (pp. 3–7). Palgrave Macmillan UK.
- Feng, D. (William). (2023). *Multimodal Chinese discourse: Understanding communication and society in contemporary China*. Routledge.
- Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge.
- Francesconi, S. (2011). Images and writing in tourist brochures. *Journal of Tourism and Cultural Change*, 9(4), 341–356. <https://doi.org/10.1080/14766825.2011.634914>
- Fu, X. (2025). 杭州“AI 创新之城”崛起的创新启示 (*The rise of Hangzhou's 'AI Innovation City': An inspiration for innovation*). Guangming Online. https://tech.gmw.cn/2025-02/23/content_37865422.htm
- González, V., Chen, C., & Sanchez, C. (2001). Cultural thinking and discourse organizational patterns influencing writing skills in a Chinese English-as-a-Foreign-Language (EFL) learner. *Bilingual Research Journal*, 25(4), 627–652. <https://doi.org/10.1080/15235882.2001.11074470>
- Guan, Y., & Forceville, C. (2020). Making cross-cultural meaning in five Chinese promotion clips: Metonymies and metaphors. *Intercultural Pragmatics*, 17(2), 123–149. <https://doi.org/10.1515/ip-2020-0007>
- Hall, E. T. (1976). *Beyond culture*. Doubleday.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.
- Hangzhou Municipal Bureau of Culture, Radio, TV and Tourism. (2023a, September 5). *Home [Chinese version]*. <https://wgly.hangzhou.gov.cn/cn/>
- Hangzhou Municipal Bureau of Culture, Radio, TV and Tourism. (2023b, September 5). *Home [English version]*. <https://wgly.hangzhou.gov.cn/en/>
- Hangzhou Municipal Bureau of Culture, Radio, TV, and Tourism. (2025, June 10). 喜讯!《什么是杭州》斩获2025年“诗画浙江”优秀文旅纪录片20强 (*Documentary What Is Hangzhou Ranks among Top 20 in 2025 "Picturesque Zhejiang" Outstanding Cultural Tourism Documentary Awards*). Hangzhou Municipal Bureau of Culture, Radio, TV, and Tourism. https://wgly.hangzhou.gov.cn/art/2025/6/10/art_1692916_58963172.html
- Hogg, G., Liao, M.-H., & O'Gorman, K. (2014). Reading between the lines: Multidimensional translation in tourism consumption. *Tourism Management*, 42, 157–164. <http://dx.doi.org/10.1016/j.tourman.2013.10.005>
- Hu, C., Zhao, Z., & Lu, C. (2024). Metadiscursive nouns in corporate communication: A cross-cultural study of CEO letters in the US and Chinese corporate social responsibility reports. *English for Specific Purposes*, 76, 28–40. <https://doi.org/10.1016/j.esp.2024.06.001>
- Institute for Planets. (2023, September 6). 5000年, 打造一座人间天堂! (*5000 years to build a Living Paradise*). RedNote. <https://www.xiaohongshu.com/explore/64f6fb5c000000001d014022>



- Jasinskaja, K. (2019). Information structure in Slavic. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642670.013.25>
- Jordan, P. (2014). Nation branding: A tool for nationalism? *Journal of Baltic Studies*, 45(3), 283–303. <https://doi.org/10.1080/01629778.2013.860609>
- Keane, M., Yu, H., Zhao, E. J., & Leong, S. (2021). *China's digital presence in the Asia-Pacific: Cultural, technology and platforms*. Anthem Press.
- Keohane, R. O., & Nye, J. S. (1998). Power and interdependence in the information age. *Foreign Affairs*, 77(5), 81–94.
- Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3–4), 243–276. <https://doi.org/10.1556/aling.55.2008.3-4.2>
- Lang, L. (2024). The Asian Games and city branding in China: Multimodal critical discourse analysis of Hangzhou's promotional videos on YouTube. *Place Branding and Public Diplomacy*, 20(4), 504–516. <https://doi.org/10.1057/s41254-024-00359-0>
- Li, T., & Pan, F. (2021). Reshaping China's image: A corpus-based analysis of the English translation of Chinese political discourse. *Perspectives*, 29(3), 354–370. <https://doi.org/10.1080/0907676X.2020.1727540>
- Liao, M.-T., & Chen, C.-H. (2009). Rhetorical strategies in Chinese and English: A comparison of L1 composition textbooks. *Foreign Language Annals*, 42(4), 695–720. <https://doi.org/10.1111/j.1944-9720.2009.01050.x>
- Lu, Y. (2018). *The emphasis of utterance in Mandarin Chinese*. China Social Sciences Press.
- Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies*, 10(4), 347–355. <https://doi.org/10.1080/17405904.2013.813770>
- Machin, D., & van Leeuwen, T. (2016). Sound, music and gender in mobile games. *Gender and Language*, 10(3), 412–432. <https://doi.org/10.1558/genl.v10i3.32039>
- Matalene, C. (1985). Contrastive rhetoric: An American writing teacher in China. *College English*, 47(8), 789–808. <https://doi.org/10.2307/376613>
- Meng, X. (2020). *National image: China's communication of cultural symbols*. Springer.
- Mills, H. (2000). *Artful persuasion: How to command attention, change minds, and influence people*. AMA Publications.
- Narins, T. P., & Agnew, J. (2020). Missing from the map: Chinese exceptionalism, sovereignty regimes and the Belt Road Initiative. *Geopolitics*, 25(4), 809–837. <https://doi.org/10.1080/14650045.2019.1601082>
- Nord, C. (2018). *Translating as a purposeful activity: Functionalist approaches explained* (2nd ed.). Routledge.
- Nye, J. S. (1990). Soft power. *Foreign Policy*, 80, 153–171.

- Nye, J. S. (2004). *Soft power: The means to success in world politics*. Public Affairs.
- Pu, X. (2019). *Rebranding China: Contested status signaling in the changing global order*. Stanford University Press.
- Rawnsley, G. (2012). Approaches to soft power and public diplomacy in China and Taiwan. *The Journal of International Communication*, 18(2), 121–135. <https://doi.org/10.1080/13216597.2012.695744>
- Roig-Sanz, D., Campanella, L., & Carbó-Catalan, E. (2025). Translation as a soft power resource: Exploring the possibilities, scope, and challenges of an interdisciplinary approach. *Perspectives*, 33(1), 1–23. <https://doi.org/10.1080/0907676X.2025.2447143>
- Simons, H. W. (1986). *Persuasion: Understanding, practice, and analysis* (2nd ed). Random House.
- Taylor, J. R. (2015). The China Dream is an urban dream: Assessing the CPC's national new-type urbanization plan. *Journal of Chinese Political Science*, 20(2), 107–120. <https://doi.org/10.1007/s11366-015-9341-7>
- Tomlin, R. (1995). Focal attention, voice, and word order: An experimental, cross-linguistic study. In P. Downing & M. Noonan (Eds.), *Word Order in Discourse* (pp. 517–554). Benjamins.
- Torresi, I. (2021). *Translating promotional and advertising texts* (2nd ed.). Routledge.
- van Leeuwen, T. (1999). *Speech, music, sound*. Macmillan Education UK.
- Verstraten, P. (2009). *Film narratology* (S. Van Der Lecq, Trans.). University of Toronto Press.
- Wang, H., & Zhu, C. (2009). Discursive mediation in translation: Representation of the Harry Wu topic in the Chinese translations of *Living History*. *Meta*, 54(4), 684–703. <https://doi.org/10.7202/038899ar>
- Wang, Y., & Feng, D. (William). (2023). History, modernity, and city branding in China: A multimodal critical discourse analysis of Xi'an's promotional videos on social media. *Social Semiotics*, 33(2), 402–425. <https://doi.org/10.1080/10350330.2020.1870405>
- Wen, S. (2012). You (有) as an informational focus marker in the construction of 'you + quantitative phrase'. *Studies of the Chinese Language*, 01, 29–37.
- Wilson, R. T., Baack, D. W., & Till, B. D. (2015). Creativity, attention and the memory for brands: An outdoor advertising field study. *International Journal of Advertising*, 34(2), 232–261. <https://doi.org/10.1080/02650487.2014.996117>
- Xinhua News Agency. (2021, June 2). *Xi stresses improving China's international communication capacity*. Qiushi Online. https://en.qstheory.cn/2021-06/02/c_628963.htm?utm
- Xu, L. (2004). Manifestation of informational focus. *Lingua*, 114(3), 277–299. [https://doi.org/10.1016/S0024-3841\(03\)00031-7](https://doi.org/10.1016/S0024-3841(03)00031-7)
- Yao, Y., & Zhuo, Y. (2018). A multimodal discourse analysis of the promotional video of Hangzhou. *English Language Teaching*, 11(10), 121–131. <https://doi.org/10.5539/elt.v11n10p121>
- Zhang, J., & Jiang, L. (2005). *A dynamic study of sentence patterns in English*. Tsinghua University Press.
- Zhang, J., & Zhang, K. (1998). *Information structure: A contrastive study in English and Chinese*. Henan University Press.



Zhang W. (2023, September 5). 以水为引，阅尽五千年人间烟火——杭州新版文旅宣传片发布 (Drawn by water, telling 5000 years of life: Hangzhou releases new promo film). *Hangzhou Daily*, 8.

Zhang, X. (2011). *The transformation of political communication in China: From propaganda to hegemony* (Vol. 29). World Scientific Publishing.

Zhu, Y. (2022). China's 'new cultural diplomacy' in international broadcasting: Branding the nation through CGTN documentary. *International Journal of Cultural Policy*, 28(6), 671–683. <https://doi.org/10.1080/10286632.2021.2022651>

Zhuang, Y. (2015). 副词“还”在递进句中的偏误成因及其矫正策略——基于汉英中介语的研析 (The causes of adverbial “*hai*” in progressive clauses and its corrective strategies: A study based on Chinese-English mediational analysis). *Journal of South China Normal University (Social Science Edition)*, 04, 176–181.

Zubizarreta, M. L. (2019). Nuclear stress and information structure. In C. Féry & S. Ishihara (Eds.), *The Oxford Handbook of Information Structure* (pp. 165–184). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199642670.013.008>

Data availability statement

The data supporting the findings of this study are available within the article. The analysed promotional videos are publicly available on the official websites of the Hangzhou Municipal Bureau of Culture, Radio, TV and Tourism. Links to these sources are provided in the reference list. The screenshots used in this article are taken from publicly available videos and are used under fair use conditions for the purposes of scholarly analysis, criticism, and comment.

Disclaimer

The authors are responsible for obtaining permission to use any copyrighted material contained in their article and/or verify whether they may claim fair use.

Notes

¹ NVivo 15 is a qualitative and mixed-methods data analysis software developed by Lumivero. For more information, please see <https://lumivero.com/products/nvivo/>.

² Praat is a speech analysis and phonetics software developed by Paul Boersma, David Weenink, and Anastasia Shchupak at the University of Amsterdam. For more information, please see <https://www.fon.hum.uva.nl/praat/>.

