A Multimodal Discourse Analysis of the Promotional Video of Hangzhou

Yinyan Yao[1] & Yanfen Zhuo[1]

[1] School of English and Education, Guangdong University of Foreign Studies, China

Correspondence: Yinyan Yao, Associate Professor of School of English and Education, Guangdong University of Foreign Studies, Guangzhou, 510006, China. E-mail: sandyyaoyy@hotmail.com

Abstract

This paper analyzes a promotional video of the Chinese city of Hangzhou from the perspective of multimodal discourse analysis informed by Systemic Functional Linguistics. By drawing on Visual Grammar as well as frameworks of intersemiotic complementarity, the paper examines how various semiotic resources, namely, the visual, audio and verbal, construe meanings and how they work together to create synergy in the video. It is concluded that the deployment of various modes in this dynamic discourse contribute to constructing city images that are glorious in history, unique in culture, picturesque in landscapes, innovative in spirits, vital in city life, and beautiful in people's hearts. The video also proves to be effective in engaging and aligning the viewers, thus functioning as a vital tool to market the city. It is hoped that this paper will provide a new perspective for semiotic studies of promotional videos in China.

1. Introduction

The advent of digital media has triggered a new trend of information dissemination and communication. Over the past two decades we have seen many cities in China deploy multi-media tools such as promotional videos to communicate their unique history and culture. Compared with the traditional tourist pamphlets, brochures, posters or magazines, promotional videos present more appealing information by integrating visual, verbal and audio modes. As promotional videos mainly function to promote tourism, attract investments, and publicize unique images of a city, they have been employed extensively accompanying the city's major events as an effective tool to market the city. For example, during the 2016 G20 Summit, quite a few promotional videos about the Chinese host city Hangzhou were released to audience at home and abroad through various channels such as CCTV, BBC, CNN, Face book, WeChat, etc. These videos had gone viral and attracted millions of viewers. The multimodal nature of promotional videos is believed to be a major factor to engage and entice the audience. Although city promotional videos have intrigued researchers in China, not many have drawn on multimodal discourse analysis (MDA) as their theoretical framework. In view of this, the present study attempts to analyze the construction of city images in promotional videos from a multimodal approach.

Multimodal studies have developed since the early 1990's and the past three decades have witnessed numerous research fruits in this field. Among various approaches to study multimodality, social semiotic perspective has been adopted widely. Informed by Systemic Functional Linguistics proposed by Halliday (1978, 1994), scholars have not only focused on grammars of single modes such as visual design (Kress & van Leeuwen, 2006), sound and music (Van Leeuwen, 1999), gestures (Martinec, 2000) and so on, but also explored how different modes are co-deployed and integrated in multimodal texts (e.g. Lemke, 1998; O'Halloran, 2003, 2008; Royce, 1999, 2007). In recent years, the study of dynamic multimodal discourse such as videos and films has gained great momentum due to the complexity in their meaning-making processes. For example, O'Halloran (2004) studies the dynamics of visual semiosis in film; Baldry and Thibault (2006) explore the transcription, annotation and analysis of video texts; Lim and O'Halloran (2012) develops macro-analytical and micro-analytical techniques to transcribe and analyze a teacher-recruitment advertisement. In China, MDA has been introduced for more than a decade. At the initial stage, scholars mainly focused on theoretical exploration (e.g. Li, 2003; Hu, 2007; Zhu, 2007; Zhang, 2009) and application of MDA to study static images and discourses (e.g. Wang, 2007; Chen & Huang, 2009; Tian & Zhang, 2013). In more recent years, considerable attention has been paid to multimodal analysis of

dynamic discourses (e.g. Hong & Zhang, 2010; Zhang, 2011; Li, 2013; Yao & Chen, 2013; Geng & Chen, 2014), which include films, videos, promotional videos and TV advertisements. However, until now, the exploration of dynamic multimodal discourse is still at the infant stage and needs further research.

This study draws on multimodal analytic tools to investigate various semiotic resources in a promotional video of Hangzhou city. The selected video titled '*Hangzhou*', released during the G20 Hangzhou Summit, is an official version sponsored and produced by the Hangzhou government. The aim of this study is to examine how visual, audio and verbal modes in the promotional videos are employed to represent, construct and project meanings, as well as how intersemiotic complementarity of these modes is realized under the approach of MDA. This study also intends to investigate how the video appeals to the audience and achieve the effect of marketing the city. The questions to be explored in this study are listed as follows:

a). How are meanings construed through visual, audio and verbal resources in the promotional video *Hangzhou*?

b). How do visual, audio and verbal resources work together to construct city images and make it appealing to audience?

2. Theoretical Framework -- Multimodal Discourse Analysis

The major objective of analyzing multimodal discourse is to investigate how meanings are constructed and communicated through different modes such as verbal, visual, audio and so on. Systemic Functional Linguistics (SFL), which is proposed and developed by Halliday (1978, 1994) in studying language as social semiotic has been widely extended to account for the meaning making by various semiotic systems in multimodal discourse. According to Kress and van Leeuwen (2006), the three metafunctions in SFL can be applied to all semiotic modes and are not specific to language. Thus, in terms of analysis, MDA also follows three metafunctions, namely, the ideational, the interpersonal, and the textual. The ideational metafunction represents the experiences in the world as well as establishes the logico-semantic and interdependency relations between clauses, the interpersonal metafunction enacts social relations and the textual metafunction makes the messages in the text into a cohesive and coherent whole (Halliday, 1978, 1994). As the promotional video under examination is mainly composed of visual images, Chinese and English titles and subtitles as well as background music, we will explore how the ideational, the interpersonal and the textual meanings are construed by each of them and how they interact with each other to create multiplying meaning.

Kress and van Leeuwen (2006) put forward Visual Grammar, a framework to analyze visual images. In Visual Grammar, the three metafunctions are renamed as representational, interactive, and compositional. The representational meanings are often realized by two types of representational structures: the narrative and the conceptual. In narrative visuals, participants are connected by a vector and "represented as doing something to or for each other" (Kress & van Leeuwen, 2006). In conceptual visuals, participants are represented "in terms of their generalized and more or less stable and timeless essence" (Ibid). Narrative processes can be distinguished into action processes, reactional processes, speech process and mental process, conversion processes based on the types of vector and participants involved; the circumstances are categorized into setting, means and accompaniment; conceptual processes include classificational, analytical and symbolic processes (Kress & van Leeuwen, 2006). When analyzing visual semiotic systems, Royce (1999) introduces Visual Message Elements (VME) to classify the features of visual elements with regard to their semantic properties, which will also be applied to our analysis.

The interactive meanings of visual semiotic system are concerned with the social relations between the producer, the viewer and the object represented, which are realized by contact, social distance, attitude and modality (Kress & van Leeuwen, 2006). The presence of gaze establishes contact between the participants and the viewers on an imaginary level, while the absence of gaze indicates objective and factual information is presented; there are two kinds of images: demand and offer, with the former meaning the participant's gaze demands something from the viewer and the latter meaning the viewers are addressed indirectly and the image offers information impersonally (Ibid). The choice of social distance through camera shots suggests closeness or distance between participants and viewers in varying degrees: close shots express intimate or personal relations, medium shots indicate social relations, and long shots connote public relations (Ibid). Attitude is categorized into subjective one and objective one based on point of views; while a frontal point of view indicates the involvement of the represented participants by the image-producer, an oblique angle indicates detachment (Ibid). Power is associated with vertical angles of camera, with high angles indicating viewer power, eye levels equality, and low angles representing power (Ibid). Modality is concerned with truth value and credibility, which distinguishes high, medium and low modality. There are eight kinds of modality markers: color saturation, color differentiation, color modulation, contextualization, depth, illumination and brightness and four types of coding orientations:

technological, sensory, abstract and the common sense naturalistic (Ibid).

The compositional metafunction integrates the representational and interactive elements into a meaningful whole through three interrelated systems: information value, salience and framing (Kress & van Leeuwen, 2006). Different zones of the image such as left and right, top and bottom, center and margin are endowed with different information values; salience is realized through factors such as foreground or background placement, relative sizes, contrasts in tonal value or color, sharpness, etc.; framing devices play a critical role in connecting or disconnecting elements in the image through frame lines (Ibid).

For intersemiotic relationship of visual, verbal and audio modes, we will follow Zhang (2009) and Royce (1999). Zhang (2009) distinguishes complementary relationship and non-complementary relationship between different modes. Complementary relationship includes intensifying and non-intensifying relations, while non-complementary relationship includes blending, embedding and context interacting (Zhang, 2009). For intersemiotic complementarity between visual and verbal modes, according to Royce (1999), ideationally, it can be realized by cohesive relations such as repetition, synonymy, antonymy, hyponymy, meronymy and collocation; interpersonally, it can be realized by reinforcement of address, attitudinal congruence, and attitudinal dissonance; and compositionally, it can be examined through the aspects of informational value, salience, framing, reading path of both the visual and verbal modes.

In the following analysis, we will examine visual resources based on Visual Grammar, then background music and subtitles in terms of three metafunctions, and finally their intersemiotic complementarities.

3. Data Description

The promotional video titled '*Hangzhou*', lasting 4 minutes and 43 seconds, is selected for the present study. It is among a series of promotional videos released by the Hangzhou municipal government during the G20 Summit in 2016, for the purpose of publicizing the culture, history and modernity of Hangzhou city to viewers both at home and abroad.

According to Iedema (2001), the analysis of dynamic discourse such as videos can be divided into six levels: 1) Work as a whole; 2) Generic stage; 3) Sequence; 4) Scenes; 5) Shot; 6) Frame. In this video, the image of butterfly has been adopted as a unique cultural symbol of Hangzhou city. The flight path of the butterfly is an implicit thread which connects the following Sequences in the video:

a)   The transformation of butterfly (length: 15 seconds);

b)   A historically and culturally famous city (length: 22 seconds);

c)   A city that enjoys a good quality of life (length: 34 seconds);

d)   A poetic and picturesque city (length: 64 seconds);

e)   A city of innovation and vitality (length: 65 seconds);

f)   A city of love (length: 50 seconds);

g)   The Hangzhou G20 Summit will ignite hope for world economy (length: 33 seconds).

From the time distribution of the Sequences as listed above, those of a poetic and picturesque city, an innovative city and a city of love are elaborated and emphasized. In the following section, the present study mainly takes "Frame" as the basic unit of analyzing visual resources. Based on the theoretical frameworks above, the present study will examine how the three modes construe meanings respectively, as well as their intersemiotic relations.

4. Multimodal Discourse Analysis of the Promotional Video *Hangzhou*

4.1 Analysis of Visual Resources

4.1.1 Visual Representational Meaning

Mainly drawing on Visual Grammar (Kress & van Leeuwen, 2006), we will look at the representational meanings of visual resources from the following VMEs: participants, process types, and circumstances. The participants and circumstances, especially the settings in this video could be divided into several categories as shown in Table 1. Based on them, six themes can be identified in the video: history and culture, lifestyle, cuisine, landscape and landmarks, economy and technology and humanity.

Table 1. Classification of visual message elements

Themes	Participants	Circumstances (settings)
History and culture	Ancient army of Song Dynasty; ancient Hangzhou residents; tea picking girl; artists.	Southern Song Capital; Longjing tea gardens; opera house; the Grand Canal; Hefang Street, etc.
Lifestyle	Old man; young ladies; young lovers; tourists; local residents.	Parks; hotels; highway; city roads.
Cuisine	Chefs; customers; Dong Po Pork; Longjing Shrimp.	Restaurants; dining table
Landscape and landmarks	Tourists; young ladies	West Lake; Leifeng Pagoda; Liuhe Pagoda; Chenghuang Tower; Su Causeway; Bai Causeway; Qiantang River; The Xixi National Wetland Park; Hangzhou Bay Bridge; Tianmu Mountain; Hangzhou Lotus Stadium; Qiantang Fuxing Bridge; Broken Bridge; Long Bridge.
Economy and Technology	AI; factory workers; young innovators and entrepreneurs, construction workers.	Office blocks of Alibaba and Net Ease; high-rise buildings; assembly lines; piers; construction sites.
Humanity	The young mother who saved a baby; the cleaner; volunteers; special education teachers.	Classroom; alleys; communities.

These themes are manifested through the flight path of the butterflies, which seems to connect the past, the present, and the future of Hangzhou. At the beginning of the video, the process of the larvae transforming into a butterfly is vividly depicted; then the butterfly flies far to the sky and brings the audience back to ancient China, unfolding a picture of prosperous and busy life in Southern Song Dynasty about 800 years ago. The grand palace of the ancient capital emerges in front of the audience as the butterfly slowly glides around the eaves, indicating the long and glorious history of Hangzhou. Then the butterfly flies to several famous scenic spots and landmarks such as the West Lake, Lingyin Temple, Liuhe Pagoda and Hangzhou Lotus Stadium, showing viewers the relaxing lifestyle of local residents, traditional Hangzhou cuisines, convenient urban transportation, and the ecological environment. Hangzhou is thus depicted as a harmonious and pleasant city where people enjoy nature and a good quality of life. The camera then switches to the modern Hangzhou metropolis as the butterflies are seen to fly across the new industrial districts in urban area. The visual images such as Office blocks of Alibaba and Net Ease and high-tech AI products present a dynamic and innovative modern city to the audience. At the end of the video, a time-lapse of the cityscape lighting up at night implies the aim of Hangzhou G20 Summit: it will ignite the hope for world economy. All these visual elements listed in the table above contribute to building a unique image of Hangzhou city as one with not only a glorious past but a bright future. With long history and unique culture, it has now transformed into an innovative and modern city. In the following part, our analysis will focus on the process types.

(1) Conceptual Representations

Visual images can be further analyzed in this video from two aspects: conceptual representations and narrative representations. In terms of conceptual representations, analytical processes and symbolic processes are the primary ones in the promotional videos. Analytical process relates "participants in terms of a part–whole structure" (Kress & Van Leeuwen, 2006). Two kinds of participants are involved in this process: Carrier (the whole) and Possessive Attributes (the parts). Symbolic process is "about what a participant means or is" (Kress & Van Leeuwen, 2006). There are two types of symbolic processes: Symbolic Attributive and Symbolic Suggestive, with the former having two participants: the Carrier and the Symbolic Attribute, and the latter having only one participant the Carrier (Kress & Van Leeuwen, 2006).

In the promotional videos, the prosperity and success of a city are usually reflected through the images of

architecture and working environment. Frames 1, 2 and 3 are regarded as analytical processes which are composed of part-whole relations. In Frame 1, the magnificent palace of the ancient capital of Southern Song Dynasty portrayed a glorious past of Hangzhou city. Frames 2 and 3 then show the audience a modern and innovative city. Frames 4, 5, 6, and 7 can also be classified as analytical processes. Frame 4 is the Lingyin Temple; Frame 5 is the lotus in blossom; Frame 6 is the Broken Bridge; and Frame 7 presents the image of West Lake. All of these images construe a peaceful, poetic, and idyllic Hangzhou which is both ancient in history and modern in city construction. Frame 8 and Frame 9 can be analyzed as the symbolic suggestive processes. Frame 8 presents the metamorphosis of a butterfly, which is an analogy to the rapid change and sharp contrast of Hangzhou over the past centuries from an ancient capital to today's modern metropolis. When two butterflies flying together, they are regarded as a symbol of love and romance in Chinese culture as they refer to a young couple in an ancient legend: Liang Shanbo and Zhu Yingtai. They were separated when alive because of disapproval of marriage from parents, but after death, their spirits turned into a pair of beautiful butterflies and flew away together. The process of the butterfly breaking from the cocoon also suggests its rebirth into a new life with courage and hardship. Therefore, the image of butterfly symbolizes romantic love, innovative spirits and hope of people in Hangzhou city, which are closely, connected to the major themes of the video. Frame 9 is a close shot of the Fuxing Bridge over Qiantang River. Three symbolic meanings are encoded in this image. The name of the bridge "Fuxing" actually means revitalization, which suggests the revival of the city into a more prosperous one. Besides, Qiantang River is famous for its tide, which connotes that Hangzhou people are standing at the forefront of innovation and entrepreneurship. Lastly, the bridge is often regarded as a symbol of connection to the outside world and a symbol of communication among people. In the video, the images of other bridges such as the oldest Qiantang River Bridge and bridges in the West Lake also appear several times, all of which serve as symbols signifying Hangzhou G20 Summit brings together 20 countries from all over the world and is a bridge for Hangzhou to communicate with the outside world.
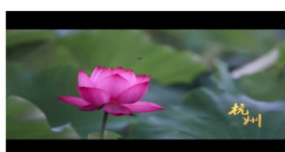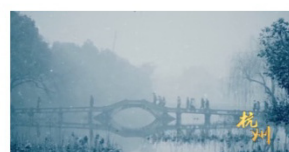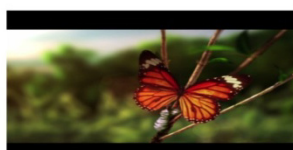
Frame 1

Frame 2

Frame 3

Frame 4

Frame 5

Frame 6

Frame 7

Frame 8
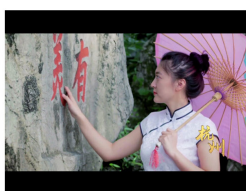
Frame 9

(2) Narrative Representations

Narrative processes in this video mainly include the action processes and reactional processes. For action processes, "the actor is the participant from which the vector emanates, or which itself, in whole or in part, forms the vector" (Kress & van Leeuwen, 2006). While non-transactional action processes have no goals, transactional processes have both actors and goals. Relational processes involve reactors and phenomena, in which "the vector is formed by an eye line, by the direction of the glance of one or more of the represented participants" (Kress & van Leeuwen, 2006).

Frames 10-15 contain action processes, either transactional or non-transactional: tourists and local residents are actors, with their actions such as playing Taiji, cycling, dining, dancing and visiting, represented in dynamic scenes. These scenes are connected in series to show the relationship between people and elements of the city. The actors include female and male, the young and the old, the Han and the minority, all of whom are ordinary

people. The landscapes, restaurants, and cuisines are the goals or circumstances of their actions. The variety of actors as well as their activities not only reflect the daily life of local people, but also demonstrate the diverse culture and vitality of the Hangzhou city. Frames 16, 17, 19 and 20 are reactional processes, with vectors formed by their eye lines. The smiles on the reactors' faces indicate happiness and satisfaction from their bottom hearts.
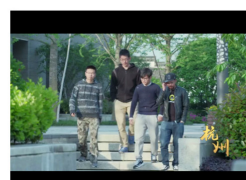

Frame 10   Frame 11   Frame 12

Frame 13   Frame14   Frame15

4.1.2 Visual Interactive Meaning

The analysis of interactive meanings of visual resources in the promotional video will focus on aspects of contact, social distance, attitude and modality.

(1) Contact

In terms of contact, in Frames 16, 17 and 19, the represented participants have direct gazes at viewers. They seem to address the viewers with a visual "you", thereby establishing an imaginary relationship with them. The presented participants also seem to demand something from the viewers. From smiles on their faces, it is clear what they demand is to invite the viewers to experience what they have experienced in the city; thus, the relationships of social affinity between them are created. The participants in Frames 18, 20, and 21 do not look at the viewers; therefore these images offer information: Frame18 depicts an actress performing the most famous Yue opera "The Butterfly Lovers"; Frame 20 describes a touching story about a young mother who was hailed as a hero after she saved a baby from falling from the high building; Frame 21 portrays a romantic scene of a young couple sitting on the famous "Bench of Love" and cuddling together while appreciating the beauty of West Lake. These three images also echo one of the major themes of this promotional video: it is a city of love.

Frame 16   Frame 17   Frame 18

Frame 19   Frame 20   Frame 21

(2) Social Distance

Social relations between the viewers and objects, buildings and landscapes can be suggested by sizes of frame (Kress & van Leeuwen 2006). The utilization of close shots, medium shots and long shots generates different social distance between the represented participants and the viewers. Frames 16, 17, 18 and 19 are close shots with only heads and shoulders of the represented participants displayed and thus generate intimate relations with the viewers. The viewers are engaged to the greatest extent through this close personal distance. The joyful

feelings of the two pretty girls in Frames 16 and 17 are revealed through their subtle facial expressions, which seem contagious to the viewers. In Frames 19 and 20, the smiling faces of the middle-aged man and the young mother who have been praised as beautiful characters with kindness also imply that Hangzhou people are helpful and warm-hearted. The close shots of them allow the viewers to come close to these public figures as if they were their friends and they were just around them. When local people living in Hangzhou are portrayed, most of the images are close shots or medium shots. For symbolic images such as ancient architecture, local cuisines, and lotus blossoms, they are depicted in close shots, thus leaving a strong visual impact on the viewers. For most of the landscapes, the camera moves from long shots to close shots. Long shots from the air afford an overview but place the viewers outside the landscape, while medium and close shots enable foregrounding objects and place the viewers imaginarily within the landscape.

(3) Attitudes

Unlike scientific and technical pictures that encode objective attitudes, the visual resources in the promotional video encode subjective ones which are done by horizontal angles and vertical angles. Horizontal angle includes frontal and oblique one, indicating producers' involvement and detachment of the represented participants respectively (Kress & van Leeuwen, 2006). Frames 22 and 23 are videoed from frontal angles. Frame 22 represents the prosperous ancient capital of Southern Song Dynasty. Frame 23 depicts Zhejiang Bridge. In both of the Frames, viewers are involved in the depicted world of the sceneries. The example of oblique angle is shown in Frame 24, which exhibits high-speed automatic machines that requires expertise for operation. Viewers watch these high-tech machines as observers and outsiders.

Frame 22   Frame 23   Frame 24

Vertical angles indicate power relationships between viewers and represented participants. Superiority, equality and inferiority towards the represented participants are realized through the high angle, eye-level angle and low angle respectively (Kress & van Leeuwen, 2006). Most of the shots are at eye level, indicating equal relationship without power difference between the represented participants and the viewers. Frames 25 and 26 are shots from a high angle as the represented participants are overlooked by interactive participants from high above. In the promotional videos, high angles are often employed to present the magnificent bird view of ancient architecture (as shown in Frame 25) and a panoramic view of the unique beauty of landscapes (as shown in Frame 26). Viewers have symbolic power over these landscapes as they are at viewers' command. On the other hand, when participants are shot from a low camera angle, they usually look strong and powerful (as shown in Frame 27). In this video, the flight path of the butterflies determine the angle of most of the landscape shots: when the butterflies fly high, the panoramic views of the Hangzhou city are shown; when they fly at the lower level, viewers get a closer look of the represented participants or circumstances.

Frame 25   Frame 26   Frame 27

(4) Modality

Modality is realized by a complex interplay of visual cues and the overall assessment is derived by the viewer (Kress & van Leeuwen, 2006). Most of the images in the promotional video are rendered with naturalistic faithfulness, in other words, high modality through the use of fully saturated, fully modulated and diversified

range of colors, full articulated and detailed background, and high degree of representation of pictorial detail, deep perspective, illumination and brightness. Besides, sensory coding orientation is adopted, as the pleasure principle is dominant in the promotional video which functions to attract viewers sensually and emotively.

4.1.3 Visual Compositional Meaning

The compositional meaning is achieved through three interrelated systems: information value, salience and framing (Kress & van Leeuwen, 2006). However, dynamic discourse is different from static images. As Baldry and Thibault (2006) point out, in progressive pictures, left and right structuring is not really useful; the New information is construed by dynamically salient informational variants or transformations while the Given is constituted by informational invariants. Thus, in the promotional video, compositional meaning cannot be analyzed in the form of static frame alone, as the images are progressive and keep changing. Each shot is an inseparable unit of the scene to generate new information to construe visual meanings as a whole. In this study, we will take one of the scenes which expound the tea culture of Hangzhou as an example to analyze the compositional meaning in the video. Visual images from Frames 28 to 33 as representatives of shots present the unique tea culture of Hangzhou. In Frame 28, a young lady in traditional costume occupies the salient position, depicted as picking tea leaves, with tea plantation as the setting; then Frame 29 shows tea leaves and the girl's hands with a close shot. With such a transition, it is easy to see the tea leaves and hands appear as new information, aiming to engage viewers with intimate relations. As the visual images unfold in time, they display constantly varying new information. For instance, in Frames 31 and 33, the main participants such as the glass of tea and the young lady drinking tea are put in the foreground, and occupy the central part of the visual space. The glass of tea in Frame 31 becomes the Given information in Frame 33.

Frame 28

Frame 29

Frame 30

Frame 31

Frame 32

Frame 33

4.2 Analysis of Audio Resource, Verbal Resource and Intersemiotic Complementarity

4.2.1 Audio Resource and Intersemiotic Complementarity

The audio mode of the promotional video mainly contains the background music, which is electro acoustic synthesis, an integration of melodies played by traditional Chinese musical instruments such as Pipa (a Chinese four-stringed lute), Chinese zither, Yue Hu, Ruan (a plucked stringed instrument), long and short bamboo flutes, etc. and western musical instruments such as violin and orchestra instruments. Local tunes with Hangzhou style, traditional Yue opera, symphony and violin concerto take turns along with the unfolding images. The choice of such background music carries strong ideational meanings. As part of traditional Hangzhou culture, local tunes, Yue opera and exquisite music played by Chinese musical instruments are shown to the viewers. Two stanzas of the melodies themselves, which can be identified as *Liang Zhu* played by the Chinese zither and the violin respectively, also reflect part of unique Hangzhou culture, as *Liang Zhu* is a very famous romantic love story in history (A Chinese equivalent of *Romeo and Juliet*). Symbolically, the music signifies that Hangzhou is not only a historical and cultural city but also a modern and internationalized one.

In terms of interpersonal meanings, the music plays a significant role in stirring the viewers' emotions and engaging them with enjoyment through tempos and rhythms, which are sometimes soothing and gentle, sometimes cheerful and lively, and other times magnificent and thrilling. Besides, the choice of classic traditional music of *Liang Zhu* can not only get closer to viewers but also arouse their empathy through its familiar melodies. Lastly, the video is without voice-over and only a few subtitles, which means the viewers can be fully immersed in the video and exert their imagination to interpret and enjoy what they see and what they hear.

In terms of compositional meanings, different instruments and melodies are played to suit for different themes of the video; thus, music serves as the key element to distinguish the transitions of themes. The whole music can be divided into seven stanzas. When the metamorphosis of butterfly is depicted, the music is the soft melody of *Liang Zhu* played by the Chinese zither. Then it switches to magnificent symphony when the butterfly flies into the Southern Song Dynasty, which echoes the theme: Hangzhou is a city with a glorious history. In the theme of presenting Hangzhou as the city with a good quality of life, the music then changes into cheerful and delightful rhythms through the Pipa. When depicting Hangzhou is a poetic and picturesque city, the video adopts local tunes as well as slow and gentle music played by Chinese zither, Pipa, flutes and so on. As images move to those of innovative and vital city, symphony with quick tempo and thrilling rhythms accompanies. When the visual images portray Hangzhou as a city of love, the music is represented by Yue opera and the violin version of *Liang Zhu*, causing emotional resonance with viewers. Finally, strong symphony is again deployed to illustrate that Hangzhou G20 Summit ignites hope for world economy.

In terms of intersemiotic relations between visual and audio modes, the employment of the background music complements the visual representations, as the audio mode supplements the visual mode to express the overall meanings of the video based on the above analysis. According to Zhang (2009), complementary relations are distinguished into intensifying and non-intensifying ones, with the former including three categories: highlighting, primary-secondary and extending and the latter including coordinating, associating and alternating. For most of the images, the audio mode intensifies the visual one, which is the primary mode. However, when the Yue opera is displayed, both the visual and audio modes are necessary, in other words, they coordinate with each other to create the integral meaning and demonstrate visually and aurally what the Yue opera really is.

4.2.2 Verbal Resource and Intersemiotic Complementarity

The language in the video includes the title of the video '*Hangzhou*' in both English and Chinese languages at the very beginning and in the end. Except for the first Sequence of the butterfly transformation, there are subtitles of both English and Chinese for the rest of the Sequences, mainly functioning to introduce the themes without more details. While the title of the video is a necessary element as the macro-theme, thus complementing the visual and audio modes through coordination, the subtitles are optional, serving to intensify the visual mode.

In terms of image-verbal relations, we will follow Royce (1999). Altogether there are seventeen clauses. Ideationally, only two processes are involved: relational and material, with 9 and 8 instances respectively, while in the visual images, these processes correspond to conceptual and narrative representations. Different from visual images, not many details are provided in the subtitles. They mainly complement the visual images by summarizing, highlighting and extending meanings through cohesive relations between VMEs and lexical items such as hyponymy, synonymy and repetition. Anchoring the visual images through subtitles, image-producers can deliver more lucid messages to the viewers. Interpersonally, the clauses are all statements offering information to the viewers, which are the same with the visual images they accompany as no direct gaze is present. Thus, there is a reinforcement of address between the visual and verbal modes. Besides, the clauses are loaded with highly positive evaluative lexis, such as "poetic and picturesque", "innovation and vitality", "beautiful", "kindness" and so on. Together with highly sensual and emotive visual images with high modality, subtitles enhance the overall interpersonal meaning of the video and attitudinal congruence is achieved. Compositionally, the titles of the city with big fonts are put in the center of the frame at the very beginning and in the end with little or no background, occupying the most salient position. Besides, at the bottom right corner, the position in which the message is regarded as New, the two Chinese characters of Hangzhou are displayed throughout the video; thus, the city name is deeply engraved in viewers' mind. The subtitles with small fonts are positioned horizontally at the bottom left, and thus are regarded as Given and Real information, while visual images take up the whole frame and gain the most visual weight.

5. Conclusion

In this paper, we draw on Visual Grammar (Kress & van Leeuwen, 2006) as well as Zhang's (2009) and Royce's (1999) framework of intersemiotic complementarity to examine how the visual, audio and verbal modes construe meanings and how they work together to create synergy in the promotional video of *Hangzhou*. From the above multimodal analysis, it can be concluded that the deployment of various modes in this dynamic discourse contributes to constructing city images that are glorious in history, unique in culture, picturesque in landscapes, innovative in spirits, vital in city life, and beautiful in people's heart. The video also proves to be effective in engaging and aligning the viewers, thus functioning as a vital tool to market the city.

The visual images in the video vividly depict the city's history and culture, lifestyle, cuisine, landscape and landmarks, economy and technology, as well as humanity through narrative processes and conceptual processes.

Two symbolic images -- butterflies and bridges are worth particularly mentioning, which appear several times in the video and become the most salient among VMEs. The butterflies carry suggestive meanings of rebirth, courage, romance, love and unique culture of Hangzhou city. Bridges mostly symbolize communication with the outside world and even the logo of the 2016 G20 Summit is composed of bridge, implying that the Summit is a bridge for international cooperation and mutual benefits in the future. Promoting the city through the city's major events has become an important means of destination marketing. The video under examination is a successful model to demonstrate how promotional video of the city can be integrated with the major events to maximize communication information and optimize communication effects. The interactive meaning of the video is realized through contact, social distance, attitude and modality. Overall, the video adopts the butterflies' perspective in which the viewers can follow their flight path and start an exciting virtual journey to witness how Hangzhou has transformed from an old identity as the ancient capital of China to a new image as a modern global metropolis. When ordinary people are portrayed, most of the images are depicted by close shots and frontal angles, thus establishing an intimate and equal relationship with the viewers. When the architecture and landscapes are presented, often long shots and high angles are adopted so as to give a panoramic view. Still, too many long shots are avoided as it is not easy to draw close interpersonal relationship through public distance. Most of the images offer information to the viewers, but viewers are also expected to participate and experience life together with the people in the video through these people's direct gazes. High modality with sensory coding orientation pleases the viewers and brings considerable pleasure to them. With regard to compositional meaning, a coherent and cohesive discourse is constructed. New information progressively appears from one shot to another, with some made to be foregrounded and salient and others backgrounded. All of them account for the reason why the viewers can be fully drawn to the video and involved to the greatest extent.

The visual mode is the primary one in the video. To be sure, success of the video is inseparable with the contribution of audio and verbal modes, in other words, background music and titles and subtitles, as they are complementary to the visual mode by reinforcing the effect and construing meanings as an integrated whole. In brief, the viewers are offered an audio-visual feast.

The present study has been able to provide a new perspective to explore the way the promotional video constructs Hangzhou's city image and its effectiveness to attract the viewers. The fact that the data has been taken from only one of the series of official promotional videos may pose a limit on the wider generalizability of the findings. Undoubtedly, multimodal analysis of dynamic discourse will contribute to a better understanding of contemporary social and cultural phenomena in China; thus, further research on this particular area needs to be carried out as an increasing number of promotional videos have been released by Chinese cities seeking to promote their images.

Acknowledgements

References

Baldry, A., & Thibault, P. J. (2006). *Multimodal Transcription and Text Analysis: A multimedia toolkit and associated on-line course book*. London: Equinox.

Chen, Y., & Huang, G. (2009). Multimodal construal of heteroglossia: Evidence from language textbook. *Technology Enhanced Foreign Language Education, 11*, 35-41.

Geng, J., & Chen, Z. (2014). The dynamic multimodal discourse analysis of documentaries. *Journal of Xi'an International Studies University, 4*, 24-28.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Arnold.

Halliday, M. A. K. (1994). *An introduction to Functional Grammar*. London: Edward Arnold.

Hu, Z. (2007). Multimodalization in social semiotics. *Language Teaching and Linguistic Studies, 1*, 1-10.

Hong, G., & Zhang, Z. (2010). An analysis of the compositional meaning of the multimodal video discourse *You and Me*. *Technology Enhanced Foreign Language Education, 6*, 20-24.

Iedema, R. (2001). Analysing Film and Television: A Social Semiotic Account of Hospital: An Unhealthy

Business. In T. Van Leeuwen, & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp.183-204). London: Sage.

Kress, G., & Van Leeuwen, T. (2006). *Reading images: the grammar of visual design*. London: Routledge. https://doi.org/10.4324/9780203619728

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin, & R. Veel (Eds.) *Reading science: Critical and functional perspective on discourse of science* (pp. 87-113). London and New York: Routledge.

Li, D. (2013). Visual grammar analysis on AD texts. *Foreign Learning Research, 2*, 7-11.

Li, Z. (2003). Social semiotics analysis on multimodal discourse. *Foreign Learning Research, 5*, 1-8.

Lim, V. F., & O'Halloran, K. L. (2012). The ideal teacher: an analysis of a teacher-recruitment advertisement. *Semiotica, 189*, 229-253. https://doi.org/10.1515/sem-2012-0032

Martinec, R. (2000). Types of processes in action. *Semiotica, 130*(3), 243-268. https://doi.org/10.1515/semi.2000.130.3-4.243

O'Halloran, K. L. (2003). Intersemiosis in mathematics and science: Grammatical metaphor and semiotic metaphor. In A. Simon-Vandenbergen, M. Taverniers, & L. Ravelli (Eds.) *Grammatical Metaphor: Views form Systemic Functional Linguistics* (pp. 337-365). Amsterdam and Philadelphia: John Benjamins. https://doi.org/10.1075/cilt.236.18oha

O'Halloran, K. L. (2004). The dynamics of visual semiosis in film. In K. L. O'Halloran (Ed.), *Multimodal discourse analysis* (pp. 109-130). London & New York: Continuum.

O'Halloran, K. L. (2008). Systemic functional-multimodal discourse analysis (SF-MDA): Construing ideational meaning using language and visual imagery. *Visual Communication, 7*(4), 443-475. https://doi.org/10.1177/1470357208096210

Royce, T. D. (1999). *Visual-verbal intersemiotic complementarity in the Economist magazine*. The University of Reading. Retrieved from http://www.isfla.org/Systemics/Print/Theses/RoyceThesis/.

Royce, T. D. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce, & W. L. Bowcher (Eds.), *New directions in the analysis of multimodal discourse*. New Jersey: Lawrence Erlbaum Associates.

Tian, H., & Zhang, X. (2013). Meanings in images and ideologies of media: A multimodal discourse analysis. *Foreign Language Research, 2*, 1-6.

Van Leeuwen, T. (1999). *Speech, music, sound*. London: Macmillan. https://doi.org/10.1007/978-1-349-27700-1

Wang, H. (2007). A multimodal discourse analysis of E. E. Cummings's poetry. *Foreign Language Education, 5*, 22-26.

Yao, Y., & Chen, X. (2013). An analysis of multimodal video discourses: Taking a corporate television advertisement as an example. *Foreign Language & Literature, 1*, 86-91.

Zhang, D. (2009). On a synthetic theoretical framework for multimodal discourse analysis. *Foreign Language in China, 4*, 24-30.

Zhang, D. (2011). On the survey of different modes in dynamic multimodal discourse: Taking TV weather forecast multimodal discourse as an example. *Shandong Foreign Language Teaching Journal, 5*, 9-16.

Zhu, Y. (2007). Theory and methodology of multimodal discourse analysis. *Foreign Language Research, 5*, 82-86.