

Exploring the Multimodal Adaptation Model of Chinese Stories from a Cross-Cultural Perspective: A Case Study of the MV “The Eternal Princess” Adapted from Cantonese Opera

Yulian Chen

School of English Language and Culture, South China Business College, Guangdong University of Foreign Studies

Author Note

The author declares no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Yulian Chen, School of English Language and Culture, South China Business College, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China. Email: yulianchen2003@163.com

Abstract

This study takes the music video (MV) of the English song “The Eternal Princess,” adapted from the excerpt *Xiang Yao* of the Cantonese opera *The Princess Chang Ping*, as a case study to explore the multimodal adaptation model of Chinese stories from a cross-cultural perspective. Through analyzing the interactive relationship between visual and auditory modes, narrative structure, and communication strategies within the MV, the research finds that the audiovisual mode serves as an effective pathway for narrating Chinese stories well. It can transcend cultural boundaries, enhancing the emotional impact and appeal of the narrative. During the adaptation process, emphasis should be placed on character portrayal and the selection of cultural elements. By drawing upon multimodal discourse analysis models, the study reveals the narrative patterns and cross-cultural significance of adapted works. This approach seeks to enable Chinese stories to reach global audiences more effectively and present a credible, appealing, and respectable image of China.

Keywords: Chinese stories, cross-cultural communication, multimodal discourse analysis, Cantonese opera, music video (MV)

1. Introduction

Drama has long been a major form of storytelling for the Chinese people. Cantonese opera repertoires contain numerous touching Chinese stories of profound cultural connotations

Received 21 July 2026; revised 27 August 2026; accepted 15 September 2026; available online 25 September 2026; version of record 30 September 2026.

Citation: Chen, Y. L. (2026). Exploring the multimodal adaptation model of Chinese stories from a cross-cultural perspective: A case study of the MV “The Eternal Princess” adapted from Cantonese Opera. *Journal of Language*, 2(3), 140–164. <https://doi.org/10.64699/26YKFG8340>

that depict “credible, lovable, and respectable Chinese images.” As an important part of Lingnan culture, Cantonese opera has been passed down for over half a century, integrating characteristics of various Chinese folk operas. In 2009, it was inscribed on the UNESCO Representative List of the Intangible Cultural Heritage of Humanity through a joint application by Guangdong, Hong Kong, and Macao.

The Princess Chang Ping, a classic Cantonese opera, is based on the story of Princess Chang Ping recorded in history of the Ming Dynasty. It was adapted by the renowned Cantonese opera master Tang Disheng from the Kunqu opera *Legend of Princess Chang Ping*. The story mainly tells the tragic love between Princess Chang Ping and her husband Zhou Shixian. In the opera, after the Ming army’s defeat, Emperor Chongzhen committed suicide by hanging himself. Princess Chang Ping narrowly escaped death and disguised herself as the deceased nun Huiqing and took refuge in a nunnery. Zhou Shixian, who survived the chaos of war, accidentally discovered the princess’s whereabouts. After several tests, the two finally recognized each other. However, their existence was eventually known to the Qing emperor. To protect their families, Princess Chang Ping and Zhou Shixian chose to bow to each other under the twin trees in the Qianqing Palace and then committed suicide by drinking arsenic to martyr themselves for their country. The story interweaves the Chinese people’s feelings for family and country with their yearning for loyal and steadfast love. *Xiang Yao* is a key selection from *The Princess Chang Ping*, depicting the final moments of Princess Chang Ping and Zhou Shixian. Its melody is adapted from “Autumn Thoughts at the Dressing Table,” a section of the ancient pipa piece “Ballad of the Frontier.” The tune is beautiful yet sorrowful, and the lyrics of the Cantonese opera are basically arranged in heptasyllabic meters, rhyming with the Shang

and Yangping tones. In 1970, singer Sham Nam Ling performed the English song “The Eternal Princess”—adapted from the *Xiang Yao* selection—in Las Vegas’ Caesars Palace to promote Cantonese opera culture. The music video (MV) was later released in 1980 and has since garnered over 10,000 views on both Bilibili and YouTube.

The Princess Chang Ping was adapted from Kunqu opera, which in turn was adapted from a historical story. This shows that adaptation, to a certain extent, determines the communication and development path of a story. However, unlike general audio-visual translation, adapted works involve more re-creation, often entailing a transformation between different art forms. Consequently, while retaining the core message of the original work, changes occur in symbolic representation, narrative techniques, and even the expression of underlying ideologies.

In light of this, this study employs Zhang Delu’s dynamic multimodal discourse analysis model (2011) and Tseng’s multimodal discourse coherence model (2013) to analyze the MV above-mentioned, combined with Hutcheon et al.’s adaptation theory (2013) to explore its adaptation strategies and effects. By analyzing the interaction between visual and auditory modalities, narrative structure, and communication strategies in the MV, it aims to investigate how Chinese stories transcend cultural boundaries and are understood and accepted by audiences from diverse cultural backgrounds, thus seeking to provide a theoretical model for the cross-cultural adaptation of Chinese stories and offer feasible approaches for the cross-cultural dissemination of such narratives.

2. Multimodal Discourse Analysis of the MV of the English Song “The Eternal Princess”

The dynamic multimodal discourse analysis model emphasizes that, in the context of

achieving communicative goals in specific situational contexts, different modalities and different symbols within the same modality interact to produce communicative meaning, and this interaction changes dynamically over time (Zhang, 2011). Thus, a meaningful piece of multimodal discourse must contain at least two vector markers: time and the units of discourse at various stages from beginning to end. On the time vector, the auditory and visual modalities do not develop in complete synchrony, which dictates that only symbols existing simultaneously within the audiovisual frame can constitute multimodal discourse. Another requirement is the presence of discourse markers, which are directional and non-linear structural unit markers. When analyzing traffic accident videos, the discourse structure ranges from larger to smaller units: figure, schema, event, episode, and story. In contrast, when analyzing weather forecasts, the structure covers image, event, episode, step, and text. This demonstrates that discourse markers depend on the communicative purpose. If a traffic accident video is treated as news reporting, the markers on the discourse vector must be determined according to the communicative goals pursued by the news, and the coordination of audiovisual symbols must present the narrative style of the news. The multimodal coherence theory proposed by Tseng (2013), through its interpretation of chains of re-occurrences, can reveal the narrative patterns of multimodal works. This theory was initially used to analyze film narratives, emphasizing that the recurrence of audiovisual features such as images, sounds, language, written text, camera movement, framing, and color drives the development of the narrative. Through the combination of recurring elements, it creates interpretable multimodal texts.

2.1 The Multimodal Discourse Structure of the MV

Since the advent of Music Video (MV) in the early 1980s, it has attracted significant

attention from both the music industry and the business sector, and been defined as “an artistic form that takes musical themes and structures as the foundation and framework for creation, uses music and images as the carriers for presenting themes, and achieves audio-visual connections through the technique of audio-visual montage” (Wang, 2013). Therefore, within multimodal discourse, MV itself constitutes a type of discourse form.

Through a comprehensive analysis of MV using ELAN, its discourse structure is divided into four hierarchical units: discourse, phase, step and shot. As a complete text in multimodal form, discourse possesses stylistic characteristics, specifically referring to the features that an MV, as a genre, must contain. The manifestation of these characteristics is the phase. A complete MV must include at least two components: publication information and music, thus consisting of at least two phases. The musical theme of an MV is primarily reflected in the music phase, while steps represent the segmentation of the musical structure. The publication information phase is also subdivided into steps and shots, with the core theme being the presentation of publication information. The division of phases and steps may vary across MVs with different designs; however, these four units adhere to the Strict Layer Hypothesis (Kaisse, 1985). Units at higher levels can share similar properties with those at lower levels but cannot equate to or fall below the level of lower-level units. Each unit comprises two modalities: audio and visual, consisting of both linguistic symbols and non-linguistic elements. The representative symbols of the visual modality include text, characters, props and colors. Among these, characters also encompass clothing, facial expressions, and body language. The representative symbols of the auditory modality include language (lyrics), singing voices, melodies, and instrumental sounds, all of which belong to the musical component throughout

the entire MV. Music carries independent meaning, yet the narrative of an MV is co-constructed by music and the representative symbols of the visual modality.

Under this context, the MV can be divided into four phases: the opening copyright information, two repetitions of the song in the middle, and the ending. Both phases and the overall discourse have independent narratives. The narrative of the discourse is composed of phases, the narrative of each phase is composed of steps, and steps are formed by consecutive shots. This concept references the method for constructing film discourse units---sequences, proposed by Janney (2010). The shot is the smallest unit, within which the visual focus remains unchanged. A shot with a fixed camera position and focal length is a static shot, while a shot with a moving camera position or changing focal length is a dynamic shot. The discourse structure of the MV is illustrated in the Table below:

Table 1

The Discourse Structure of the MV

Discourse	Phrases	Steps	Shots
MV	Intro	Copyright information	1-3
		Rest of the Intro	4-6
	The First Repetition	wind blows flowers to the sky	7-8
		dark moon hanging way up high	9-10
		Let's pray our parents before we die	11-12
		though we have no time to say good bye	13-14
		tell me why you don't smile	15
		don't just knelt down/my princess please don't cry	16-17
		for our true love just like a dream/ let us fly to the moon	18-19
		side by side	20-22
		The Outro	23
	The Second Repetition	The Intro	24-25
		wind blows flowers to the sky	26-27
		dark moon hanging way up high	28-29
		Let's pray our parents before we die	30-31
		though we have no time to say good bye	32-33
		tell me why you don't smile	34
		don't just knelt down/my princess please don't cry	35-36
		for our true love just like a dream/ let us fly to the moon	37-38
	The Ending	side by side	39-41
		The Ending	42-43

The musical portion of the MV consists of an intro, two repetitions of the song, and an outro. The first repetition of the song has an intro, and the second repetition has an outro; aside from this, the two song sections are entirely identical. Therefore, the intro and outro are categorized into different stages. The intro stage comprises two steps: the first half includes shots 1 to 3, which primarily convey copyright-related information, hence it is separated from the latter half of the intro. The singing stage is segmented primarily based on the lyrical information combined with the corresponding shots. For instance, in the first singing stage, shots 16 and 17 belong to the same step, containing the two lines: “Don’t just knelt down. My princess please don’t cry.” This is because “don’t just” appears in shot 16, while “knelt down” and “my princess please don’t cry” are in shot 17. To ensure that one step conveys a complete lyrical message, these two lines are grouped into a single step. In the MV, several consecutive shots represent one sentence of the song, making them the smallest unit.

2.2 Repetition Chains in the MV

According to Tseng’s (2013) analysis of film narrative, the smallest units constituting multimodal discourse repetition chains are individual elements within a film, such as characters, objects, and settings. These elements are presented, tracked, and re-presented through different modalities in the film, thereby forming coherent identity chains. These identity chains aid the audience in understanding the film’s narrative structure and reveal deeper socio-cultural meanings. Using ELAN to analyze the shots of the MV, the recurring auditory elements include the song and the melody, while the visual elements include characters, scenes, props, and colors. Some of these elements are both recurring individual elements and part of recurring composite elements. These recurring elements are listed in Table 2.

As can be seen, the MV features a structural repetition, with the second and third stages containing a repeated sequence of 17 shots. The singer appears only during the vocal segments, either in half-body or full-body shots, consistently as the focal point of the shot. The singer is dressed in a black suit, and their body movements are coordinated with the singing.

There are a total of 16 shots featuring the princess. Among these, except for one isolated shot in the intro that transitions from a hand close-up to a facial close-up, all other shots show her sharing the frame with other characters. In shots where she appears with the singer, the princess serves as a background figure. In shots with the consort, she is the focal point, engaging in interactions such as toasting, exchanging glances, and embracing. The consort has three separate shots, depicting the actions of drinking, turning his head, and looking directly forward. These three shots all occur during the singing segments. The costumes and styling of the consort and princess closely resemble those of the characters in the Cantonese opera *The Princess Chang Ping*.

Fifteen shots in the MV clearly depict the setting. Among them, eleven shots show a room with classical Chinese characteristics, while four are close-ups of props, with the wine cup recurring multiple times. The overall color palette of the MV leans towards dark tones, dominated by red, black, gray, and blue. Only six shots do not employ dominant color rendering techniques. In the remaining shots, the primary dominant colors used are gold and gray, with smaller amounts of black and blue. These four colors are mostly applied alternately, with a few shots combining them.

Table 2

Repetition Chains in the MV

Shot	Music		Singer	The Princess	The Prince Consort	Scene	Prop	Base Color
1	Melody			Reflection			Wine Vessel (Close-up)	Blue
2	Melody			Drinking Together (Long Shot)			Round Window (Close-up)	Black
3	Melody			Drinking Together (Close Shot)				Gold
4	Melody			Hand Pouring Wine (Close-up)			Wine Pot & Wine Vessel	Blue
5	Melody			Face (Close-up)				
6	Melody						Wine Vessel (Close-up)	Blue
7		Song	Full Body	Background		Appear		Gold
8		Song			Drinking			
9		Song	Half Body		Background	Appear		Gold
10		Song		Face (Close-up)				
11		Song	Head					Black
12		Song		Dynamic (Princess → Princess & Consort)				Gold-Gray
13		Song	Full Body		Background	Appear		Gold
14		Song			Head Turn			Gray
15		Song	Half Body	Background		Appear		Gold
16		Song		Sitting (Long Shot)		Appear	Appear	Gray
17		Song	Sitting Pose	Background (Face Close-up)		Appear		Gold
18		Song	Head			Appear	Background	Gray
19		Song		Dynamic (Princess → Consort) (Close-up)		Appear	Appear	Gray-Gold
20		Song	Half Body			Appear	Background	Gray
21		Song		Face (Medium Close-up)				Gold-Gray
22		Song			Face (Close-up)			Gold
23	Melody			Sitting (Long Shot)		Appear	Background	Gray
24-41	Repeat 5-22							
42	Melody			Sitting (Long Shot)		Appear		Gray
43	Melody			Embrace				Gray-Dark

3. Multimodal Discourse Analysis of Cross-Cultural Adaptation

Htutcheon et al. (2013) argue that understanding an adapted work requires attention not only to the work itself but also to the purpose and context of the adaptation, which are crucial factors for comprehending the adapted piece. The MV of the English song “The Eternal Princess” is adapted from *Xiang Yao*, a selected excerpt from the Cantonese opera *The Princess*

Chang Ping. This constitutes a cross-cultural adaptation of traditional Chinese opera, with purposes extending beyond artistic appreciation to cultural communication. When Sham Nam Ling first performed the English song in the United States, her aim was to introduce Cantonese opera to the audience. When this MV was produced in 1980, the art form of MV had just emerged, and both hardware and software technologies were in their infancy. Therefore, conveying complex cultural characteristics through a simple format is a key feature of this MV. The following analysis will examine the cross-cultural communication intentions behind its adaptation by conducting a multimodal analysis of the MV.

The full Cantonese opera *The Princess Chang Ping* is divided into six acts: *Tree Oath*, *The Fragrant Calamity*, *Begging for the Corpse*, *Encounter at the Nunnery*, *Submitting the Memorial*, and *Xiang Yao*. As the sixth and final act, *Xiang Yao* represents the climax of the story and the peak of dramatic conflict. This segment best embodies the core values of traditional Chinese culture—faithful love and patriotic devotion to the nation. From an artistic perspective, its soundtrack is adapted from *Autumn Thoughts at the Dressing Table*, a classic piece of ancient Chinese music that serves as a typical representative of traditional Chinese melodies. Additionally, the performance, centered on the theme of marriage, builds a cultural bridge with Chinese characteristics, using a distinctive form to convey a universal ceremonial meaning.

The adapted MV features a complete narrative structure with an independent theme. Its discourse consists of four stages, with the second and third stages being identical repetitions. This repetition results in two renditions of the song and two retellings of the story, allowing the audience to go through the understanding process twice. From a cognitive perspective, the

audience can acquire more information through these two identical narrations while bearing a relatively light cognitive load. As a form of stage drama, Cantonese opera primarily relies on actors' performances. In the *Xiang Yao* excerpt, the princess's singing lasts 1.26 minutes, during which she performs as many as 19 body movements. Before the prince consort begins singing, he has 9 body movements captured on camera; after he starts singing, there are 6 interactive body movements between him and the princess. By contrast, the MV's primary form of expression is the song itself. The two performers have fewer than 5 body movements in total, and the singer's gestures are no more than simple actions like raising the head or lifting the hands to match the singing rhythm. The relationship between the MV's visuals and the song is complementary: together, they construct the meaning of a multimodal discourse. This meaning differs from the message conveyed solely by the lyrics, thereby enriching the content of cross-cultural communication.

3.1 Cross-Cultural Significance of the Auditory Mode

The libretto of Cantonese opera follows a structure of seven characters per line, with almost every line rhyming. The English lyrics similarly maintain a rhythm of 7 to 8 syllables per line, with end rhymes. For example, the Cantonese libretto uses front repetition in the line “偷偷看，偷偷望” (secretly glance, secretly look); in the English version, this segment is rendered as “tell me why you don't smile,” which employs end rhymes for phrases. The rhythm of the English lyrics maximally preserves the rhythmic features of the Cantonese opera libretto—this aligns with the primary principle of music translation, which emphasizes “fidelity to the melody” (Low, 2008). Although the soundtrack was simplified in the MV, it still retains the *suona* (a traditional Chinese wind instrument) and gongs and drums, iconic instruments of

Cantonese opera.

The libretto of *Xiang Yao* in Cantonese opera is a duet between the princess and the prince consort. However, the English song is a solo. The narrative mode was adapted from a “question-and-answer” format to a one-sided expression of the prince consort’s innermost thoughts. The story’s background, characters, events, and ending are all presented from the prince consort’s perspective. The first line of the English lyrics is a translation of the first line of the original libretto, setting the time and context. Subsequent lines shift to a first-person narration of the events. Key words describing characters and their relationships include “us,” “parents,” “you” and “princess”; verbs and verb phrases depicting actions include “pray,” “die,” “say good-bye,” “knelt down” and “cry.” These key terms outline the story of a royal couple who sacrifice their lives to follow their parents (in death). In terms of emotional expression, the English lyrics adopt two approaches: direct emotional outpouring, such as “don’t smile,” “cry” and “true love”; metaphors, such as “like a dream” and “let us fly to the moon side by side.” By contrast, the Cantonese opera libretto uses more modal verbs to convey emotions, such as “fear,” “wish,” and “long for.” Overall, the narrative of the English lyrics is more direct and straightforward.

From a cross-cultural perspective, the English lyrics convey the information of a tragic story, the rhythm preserves the rhythmic features of the Cantonese opera libretto, and the musical melody is retained. It can be said that the song itself has achieved cross-cultural communication to a certain extent. However, when it comes to “conveying the story of China” and “presenting a credible, adorable, and respectable image of China,” the MV’s visual mode plays a complementary role, making this goal more fully realized.

3.2 Cross-Cultural Significance of the Visual Mode

According to the discourse structure of the MV, the recurrence of visuals unfolds in four phases, with the second and third phases being repetitive.

In the first phase (the prelude), the visual and auditory modes maintain a non-complementary parallel relationship. The visual mode conveys two key pieces of information: one is copyright details, and the other is the story itself.

During the second and third phases (the singing sections), the visual transitions align with the rhythmic structure of the song's segments rather than the narrative of the lyrics. Notably, the singer—who appears repeatedly—has no counterpart in the lyrical narrative. Dressed in a typical modern style, the singer appears in the MV in the role of a performer, creating a discrepancy with the “princess consort” perspective adopted in the lyrics. In the singer's shots, the princess consort and the princess are presented as background figures, and shots of the royal couple alternate with those of the singer. Through the interweaving of time-space frames and narrative perspectives, a cross-narrative is constructed, delivering an emotional experience that blends the “inside” (the story of the royal couple) and “outside” (a narrator telling “our story”). The interweaving of time and space manifests in two ways. One is the composition of focus and background: the foreground features characters of the “present,” while characters of the “past” are inserted as the background. The other is the contrast between performers, which is reflected in the contrast between the modern singer performing an English song and the actors in ancient costumes. As a result, the MV offers more narrative dimensions than the song alone. While the song only encompasses characters, emotions, and a basic story, the MV builds on this foundation by adding dimensions of “modern vs. past” and “narrator vs.

story characters.” This time-space-intertwined narrative draws the audience into a broader narrative framework, blurs cultural boundaries, and enhances the sense of engagement for cross-cultural audiences.

The recurrence of colors further strengthens this sense of engagement and sets the tone for the story. Most shots in the MV use primary color rendering, with four colors—blue, black, gray, and gold—alternating, and occasional shots combining two or more. The purpose is to create a nostalgic tragic atmosphere (Thorstenson, 2015), as these colors reflect universal cultural emotions.

The actors’ shots in the MV serve as the primary vehicle for presenting the story. Therefore, as a Chinese story for cross-cultural communication, these shots must carry both culturally specific symbols and culturally universal symbols, while also aligning with the lyrics to advance the narrative. The princess is the main performer of the dramatic conflict, and the princess consort’s performance mostly complements hers. The MV opens with a shot of the princess’s reflection in a wine vessel, followed by a transition from a long shot to a close-up of her toasting with the princess consort. The prelude concludes with a close-up of the princess pouring wine. In line with Tseng’s (2013) model, the princess’s appearance in the prelude is presented in a gradual, dynamic manner. Among the 15 shots in the singing sections, 8 feature the princess—7 of which are close-ups or facial close-ups, mostly depicting her in sorrow. However, in the final shot of the couple embracing, the princess’s face shows joy. Combined with the lyrics’ narrative (which spans from life to death), this shift from sorrow to joy amplifies the sense of tragedy.

Shots of the princess and princess consort together appear in every phase. In the prelude,

a shot of the couple toasting, transitions from a long shot to a close-up, which lays the groundwork for the story. In the singing sections, such shots appear corresponding to the lyrics “let’s pray our parents before we die” and “let us fly to the moon side by side.” Both are dynamic shots: one shifts from a front view of the princess to a front view of the couple (each occupying half the screen); the other shows the princess turning to look at the consort, with the shot transitioning from a close-up of the princess to a close-up of the consort. These two instances deepen the lyrics, illustrating the mood and state in which the couple undertakes the actions described in the song. While in the conclusion, there are two shots of the couple together—one is a full-body shot of them sitting on either side of a square table, and the other is a shot of them embracing—both serving to conclude the story. Shots of the royal couple together supplement and extend the song’s narrative, using visual imagery to convey the characters’ psychological activities. The expressions and movements in the actors’ performances are rooted in universally recognized cultural understandings.

3.3 Interaction Between Visual and Auditory Modes

Hutcheon (2013) argues that a significant part of the value of an adapted work stems from its relationship with the original. How well the repetition and differences between the adaptation and the original are handled determines the extent to which the audience can gain the expected emotional experience from a sense of familiarity and novelty. For cross-cultural audiences, repetition of elements from the original work becomes a source of novelty, while universal cultural elements and aspects similar to their own culture serve as sources of familiarity.

In the MV, the visual presentation of Cantonese opera elements complements the

audience's understanding of the song (the auditory mode). Firstly, the actors' costumes clearly identify the Cantonese opera origin. Secondly, the movements—most notably the wine-drinking action—play a key role. Although the English lyrics contain no reference to drinking, wine is a central thread in the original Cantonese opera: it begins with wine being used to pay tribute to the late emperor, and later, the couple drinks poisoned wine together to die for their country. Both the utensils used for drinking and the drinking gestures bear distinct characteristics of the original work.

Two key events in the story—marriage and worship—are co-created by the visual and auditory modes. The lyrics do not directly mention marriage, only using words like “let’s” and “love” to imply the couple’s relationship. The depiction of marriage is primarily conveyed through the visual mode: the actors’ costumes and their movements. In the prelude, there is a shot of the couple toasting; in the epilogue, there is a shot of them sitting opposite each other at a table with wine pots and vessels, gazing at each other. These shots reflect the characteristics of ancient Chinese wedding rituals. In the song, “pray” and “parents” indicate the worship event, while the action of “knelt down” embodies the traits of traditional Chinese worship rituals. Although there is no shot of the couple kneeling, the prelude includes a close-up of the princess pouring wine. In Chinese culture, wine is used in both worship and wedding ceremonies: pouring wine for a toast signifies marriage, while pouring wine onto the ground signifies worship. Thus, the appearance of the wine-pouring action creates two possible expectations for the audience.

In the prelude of the MV, the visual and auditory modes maintain a non-complementary relationship, which evolves dynamically from independence to interaction. Correspondingly,

the audience's attention shifts from being primarily focused on the visual mode to engaging with both modes. The first shot uses blue as the base color, featuring a wine vessel, the princess's reflection in the vessel, and copyright information. At this point, the prelude consists of *suona* (a traditional Chinese wind instrument) sounds with an indistinct melody, so cross-cultural audiences are not yet drawn to the music. However, the wine slowly flowing into the vessel and the princess's reflection introduce the story's characters and events—here, wine and the wine vessel act as symbols representing the two events of worship and marriage. Subsequently, the character's reflection becomes increasingly clear, revealing even their costumes, which to some extent stimulates the audience's understanding of the events. When transitioning to the second shot, the music introduces two gong and drum beats that mark the rhythm, gradually forming a melody. At this point, the base color of the shot shifts to black, leaving only the silhouettes of the princess and princess consort toasting inside a circular window in the center of the screen (with copyright information still retained). Although the auditory and visual modes are still transmitting independent information, they begin to interact within the contextual framework of Cantonese opera characteristics. In the next part of the prelude, the copyright information disappears, and the screen shows a close-up of the princess pouring wine—marking the start of the story. The music's rhythm gradually accelerates, accompanied by several gong and drum beats, heralding the end of the prelude. In this segment, the auditory and visual modes interact within the context of the story's introduction.

At the start of the singing sections, the lyric subtitles appear on the screen before the vocals begin. For the auditory mode, these subtitles are non-complementary redundant symbols, but they help cross-cultural audiences better understand the linguistic symbols of the song.

Each line of lyrics is paired with two alternating shots: a shot of the singer first, followed by a shot of the actors. The lyric subtitles appear before the vocals to help the audience comprehend the content in advance. The singer's performing gestures in the shots reinforce the auditory mode; when combined with the actors (who appear as the background in the same shots), they expand the story. Meanwhile, the repeated shots of the actors, together with the song, serve to advance the narrative.

3.4 Cross-Cultural Significance of Multimodal Repetition Chains and Dynamic Interaction

In the experiment by Markhabayeva & Tseng (2024), it was found that the “identification system” formed by multimodal repetition chains serves as a crucial clue for film audiences to comprehend the narrative. For cross-cultural audiences, culturally specific information often presents cognitive challenges. Therefore, a series of multimodal repetition chains that integrate both cultural commonalities and specificities can aid their understanding of the narrative. For art forms like MVs, this can also help the audience appreciate the work.

In terms of form, “The Eternal Princess” was created to promote Cantonese opera culture. Consequently, the MV must both reflect the characteristics of Cantonese opera and be accessible to cross-cultural audiences. The melody represents a cultural commonality, as do the English lyrics and the singing style of popular music. In contrast, the recurring use of suona and gong-and-drum instrumentation, along with Cantonese opera costumes, conveys the distinctive cultural features of Cantonese opera.

In the story conveyed by the lyrics, the princess's identity is explicitly presented, while the consort's identity is more implicit, requiring the audience to infer it. The acts of worship

and death are clearly depicted, while marriage must be indirectly inferred through descriptions of love. In the visual modality, the two actors repeatedly appear in royal wedding attire. Their actions of toasting, exchanging glances, sitting separately at a table, and embracing all supplement and elaborate on the story's theme of marriage. Within this repetition chain, toasting and sitting separately at the table are culturally specific elements, while using exchanged glances and embraces to express love represents a cultural commonality.

The tragic ending of the story is indirectly expressed in the lyrics through words like “die” and “cry.” However, combined with the dark color tones presented in the visual modality and the princess's sorrowful expression maintained from beginning to end (except for the final shot), it is very clearly presented to the audience. The use of color and expression to convey tragedy is a cultural commonality, but the tragic ending here reflects characteristically Chinese values. The last line of the lyrics subtly expresses an outcome of joint death without regret. This aspect might cause confusion for cross-cultural audiences, potentially prompting them to explore further and understand the traditional Chinese sentiment of devotion to both family and nation.

4. Audiovisual Adaptation Models for Chinese Stories

Adapting traditional Chinese opera, songs, and even literary works into MV form for cross-cultural dissemination is one way to tell China's stories well, communicate China's voice effectively, and present a credible, appealing, and respectable image of China. “External communication should achieve understanding and communication at the level of values to allow the world to understand authentic Chinese culture” (Yu, 2021). Therefore, adaptations aimed at cross-cultural communication should follow a “connection” and “comprehension”

model. The form and content of the “connection” should be measured by the standard of “comprehension.” Based on the analysis of the example above, an audiovisual adaptation model for Chinese stories aimed at cross-cultural communication is constructed here, based on Shannon’s model of information transmission.

In cross-cultural communication, encoding and decoding are two key links in the process of information transmission. Encoding refers to the sender converting the message into a transmittable form, while decoding refers to the receiver interpreting the received message to derive meaning. In communication processes involving different cultural specificities, the encoder needs to consider the cultural characteristics of the receiver so that the message can be decoded correctly. Assume there is an entity possessing cultural-specific attributes (C_x) and cultural commonalities (C_{common}) from culture X, which needs to be adapted so that a decoder from culture Y can interpret it correctly. This process can be expressed through the following mathematical model:

$$E'(O) = f(E(O), C_x, C_{common}, C_y)$$

The specific steps are as follows: the original information O is first encoded into $E(O)$, which may contain cultural-specific attributes C_x of culture X. To enable decoding by a decoder from culture Y, it is necessary to consider cultural commonalities C_{common} —elements or concepts shared across different cultures. Simultaneously, the cultural-specific attributes C_y of culture Y must be considered, which may include specific values, beliefs, customs, etc., and these attributes will influence the decoding process. The adaptation function f is responsible for integrating these elements to generate a new encoding $E'(O)$, which should be more easily and correctly decoded by the decoder of culture Y. This model emphasizes that

in cross-cultural communication, the encoder must carefully adapt the information to align with the decoding habits and cultural characteristics of the target culture.

When discussing the selection of storylines for novels, Mr. Lao She believed that novels are humanity's concern for itself, the self-awareness of human society, and the record of human life experiences. Therefore, stories are primarily about people and must hold value for human development (Lao, 2012). The key to telling China's stories well also lies in telling the stories of the Chinese people well and presenting a credible, appealing, and respectable image of China. Thus, both in the selection of stories and in the encoding of information, a people-oriented approach is essential. First, the selected stories should possess universality, capable of touching upon emotions and experiences shared by all humanity, i.e., possessing cultural commonalities C_{common} . These stories may revolve around universal themes such as love, courage, and sacrifice. On this foundation of commonality, the characters' expressions should have clear C_x characteristics, which could be reflected in clothing, behavior, or language, for example. In the background elements that enhance the character portrayals, C_y should be considered, which could involve supporting characters, props, or other forms of cultural metaphors. Most importantly, the support of linguistic symbols is essential.

From this perspective, compared to single text or images, the audiovisual mode can provide a richer sensory experience, including sound, music, imagery, and motion pictures, thereby enhancing emotional transmission and information interpretation. The use of popular forms of expression, as opposed to culturally specific forms, means that the story's presentation is closer to the daily experiences of the target audience, making it easier for audiences from different cultural backgrounds to understand and accept, thereby increasing the breadth and

depth of its dissemination. When considering cross-cultural adaptation, Chinese stories should not only retain their cultural characteristics but also adopt popular forms of expression to make them more comprehensible and acceptable. The audiovisual adaptation model for Chinese stories can be expressed by the following formula:

$$E'(O_{\{story\}}) = f(E(O_{\{story\}}), C_{\{x_{\{char\}}\}}, C_{\{common_{\{themes\}}\}}, \\ C_{\{y_{\{culture\}}\}}, C_{\{pop_{\{expr\}}\}})$$

During the adaptation process, the images of Chinese characters $C_{\{x_{\{char\}}\}}$ and universal themes $C_{\{common_{\{themes\}}\}}$ in the story should be carefully preserved to convey the unique charm of Chinese stories. The adaptation needs to take into account the values and acceptance of the target culture $C_{\{y_{\{culture\}}\}}$, ensuring that while maintaining the original essence of the story, it aligns with the expectations of the target audience. Incorporating popular forms of expression $C_{\{pop_{\{expr\}}\}}$ allows the story to be presented in a way that is closer to the daily lives of the target audience. Such expressions may include the use of widely recognized narrative structures, humorous elements, emotional resonance points, etc., making the story more accessible to a broad audience. The adaptation function f aims to create an adapted story that not only embodies Chinese characteristics but also resonates internationally. In this way, the adapted Chinese story $E'(O_{\{story\}})$ will have broader appeal, play a greater role in cross-cultural communication, and achieve more effective dissemination and deeper cultural understanding.

5. Conclusion

Chinese stories carry the profound weight of a five-thousand-year-old civilization and embody the spiritual core of the Chinese nation. They serve as a bridge connecting the past and

the future, as well as linking China with the world. How to tell Chinese stories in a more vivid and fascinating manner, enabling the world to better understand and appreciate them, is a subject that requires continuous exploration and practice.

Taking the MV of the English song *The Eternal Princess*, adapted from *Xiang Yao* (a selection from the Cantonese opera *The Princess Chang Ping*), as an example, this paper attempts to analyze the adaptation model of Chinese stories from a cross-cultural perspective. Through an in-depth analysis of the interactive relationship between audio-visual modalities, narrative structure, and communication strategies in the MV, it is found that the audio-visual model is an effective way to tell Chinese stories well. Rich sensory experiences and popular forms of expression can transcend cultural boundaries, enhance the story's appeal and attractiveness, and make it easier for audiences from different cultural backgrounds to understand and accept. In the adaptation process, attention should be paid to character portrayal and the selection of cultural elements. On the basis of preserving the story's core characters and cultural characteristics, integrating elements and expression methods familiar to the target cultural audience is essential to better achieve the goal of cultural communication. The dynamic interaction between modalities and the repetition chain are the means to achieve this goal. The MV work analyzed in this study is relatively outdated, and the presented modal interaction is still relatively superficial. In the future, the summarized mathematical model can be used to analyze more complex works, and more objective and quantifiable suggestions can also be put forward for the adaptation of Chinese stories in a targeted manner.

Acknowledgments

The author wishes to acknowledge the support from the project "Research on the Integration

of Lingnan Culture into the Ideological Education for English Majors: A Practical Teaching Perspective” (Project Code: 2022GXJK408) and from the Center for Multilingual Translation and Dissemination Studies of Chinese Culture (Code: 2021WZID008).

References

- Htutcheon, L. et al. (2013). *A theory of adaptation*. Routledge.
- Janney, R. W. (2010). Film discourse cohesion. In C. R. Hoffman (Ed.), *Narrative revisited* (pp. 245–265). Benjamins.
- Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology*. Academic Press.
- Lao S. (2012). 我怎样写小说[*How I write novels*]. 译林出版社[Yilin Press].
- Low, P. (2008). Translating songs that rhyme. *Perspective: Studies in Translatology*, 16(1–2), 1–20. <https://doi.org/10.1080/13670050802364437>
- Markhabayeva, A., & Tseng, S. H. (2024). The role of identification systems in narrative comprehension. *Journal of Film and Video*, (1), 1–15.
- Thorstenson, C., Pazda, A., & Elliot, A. (2015). Feeling blue and seeing blue: Sadness may impair color perception. *Psychological Science*, (9), 1502–1509. <https://doi.org/10.1177/0956797615597672>
- Tseng, C. (2013). *Cohesion in film: Tracking film elements*. Palgrave Macmillan.
- Wang A. M (2013). MV 的概念界定、分类和五个发展阶段[Conceptual definition, classification, and five development stages of MV]. 音乐传播[*Music Communication*], (2), 72–82.
- Yu G. M. & Pan J. B. (2021). 试论我国国际文化影响力传播的路径与策略[On the pathways and strategies for enhancing China’s international cultural influence]. 传媒观察[*Media Observer*], (4), 13–18.
- Zhang D. L. & Yuan Y. Y. (2011). 动态多模态话语的模态协同研究——以电视天气预报多模态语篇为例[Research on mode coordination in dynamic multimodal discourse: A case study of multimodal texts in TV weather forecasts]. 山东外语教学[*Shandong*

Foreign Language Teaching], (5), 9–16.