

Auditory Meaning-Making in Multimodal Documentary Discourse: Functions, Features, and Synergies

Wenhan Pan

School of Foreign Languages, Guangzhou College of Commerce, Guangzhou, China
panwhsabrina@outlook.com

Received 07 July 2025 | Received in revised form 20 August 2025 | Accepted 24 August 2025

APA Citation:

Pan, W. (2025). Auditory Meaning-Making in Multimodal Documentary Discourse: Functions, Features, and Synergies. *Journal of English Language Teaching and Linguistics*, 10(2), 2025, 335-354.
doi: <http://dx.doi.org/10.21462/jeltl.v10.i2.1696>

Abstract

This study explores how sound helps shape cultural meaning in the English-language documentary Festive China. While most multimodal research has focused on images and text, sound has received less attention. This paper aims to fill that gap by examining the use of narration, background music, and ambient sound in the film. The analysis focuses on five key features of sound: loudness, pitch, rhythm, timbre, and brightness. The study uses a qualitative method based on Systemic Functional Linguistics and multimodal discourse analysis. It looks at how sound works with visuals and language to build meaning in three ways: what is shown, how the audience is engaged, and how the content is organized. The results show that narration explains ideas and builds emotional tone. Background music creates mood and supports the story's rhythm. Ambient sounds add realism and connect viewers to the setting. These elements often work together to support themes like celebration, memory, and tradition. The sound design also reflects cultural values such as balance and harmony. By focusing on audio modality, this study adds to current multimodal research and shows the importance of sound in meaning-making. It also offers insights for educators who want to help students better understand how documentaries use sound to share cultural ideas.

Keywords: multimodal discourse analysis, documentary studies, Systemic Functional Linguistics, cultural meaning-making

1. Introduction

Multimodal discourse analysis (MDA) has grown quickly in recent years. Many researchers have studied areas like critical multimodal discourse (Tian & Pan, 2018), aesthetic expression across modes (Dong & Yuan, 2021), image-based metonymy (Feng, 2017), and translation in

multimodal texts (Wang, 2023). These works show that scholars are paying more attention to how language, images, sounds, and gestures work together to create meaning.

Halliday (1989) pointed out that people do not only use language to communicate. They also use pictures, sounds, and body movement. These different forms can carry meaning. Based on this idea, researchers started using MDA to study how different modes combine in communication. Kress and van Leeuwen (2006) built a visual grammar system based on systemic functional linguistics (SFL). Their model helps people understand how visuals express meaning, show relationships, and create structure. Later, Martinec and Salway (2005) developed a method to study how words and images work together.

So far, most research has looked at visual modes. But audio modes, especially in documentaries, have not received enough attention. In documentaries, sound plays a key role. It includes voice-over narration, background music, ambient noise, and tone of speech. These sounds help tell the story, shape the mood, and support meaning. Chion (2021) and van Leeuwen (2004) explained how features like pitch, timbre, rhythm, and loudness can create meaning. Still, these ideas have not been widely used in cultural studies.

As a type of multimodal media, documentaries use sound to explain ideas, show emotions, and build scenes. But many studies, especially those about Chinese culture, focus only on what is seen or said. Recent scholarship on Chinese documentaries has largely emphasized visual representation, image-text relations, or symbolic interpretation, while sound has remained marginal. For example, Li et al. (2020) provided a historical review of the evolution of Chinese documentary programs, outlining six developmental phases, but notably did not address the role of sound or auditory elements. Similarly, Nai and Hassan (2022) applied systemic functional linguistics to analyze a BBC documentary on Chinese New Year, yet their focus remained on visual-verbal interaction. Wei (2024) examined how multimodal metaphors are structured in the Sino-British documentary *Through the Seasons: China*, finding that although multiple modes are employed, sound is seldom conceptualized as a standalone semiotic mode (Wei, 2024). Collectively, these works show that the auditory dimension of meaning-making in cultural documentaries has been underexplored. This neglect creates a gap in how we understand the role of sound in meaning-making.

This study tries to fill that gap. It focuses on how sound works in the English-language documentary *Festive China*. The film introduces Chinese traditional festivals and customs. It gives rich examples of how narration, music, and ambient sound work with images and words to create cultural meaning.

This research asks four questions:

1. What kinds of sounds are used in *Festive China*, and what are their main features?
2. How do these sounds help build cultural meaning?
3. How do sound, image, and text work together to create a full experience?
4. What strategies are used to manage sound so that it fits the story and supports the mood?

To answer these questions, the study uses a qualitative method. It looks closely at how sound is designed and how it works with visuals and narration. The goal is to better understand how documentaries use sound in cultural storytelling. At the same time, this research gives ideas for how to teach students to understand and talk about sound in multimodal media.

2. Related Work

This chapter reviews past research on sound in documentary discourse. It focuses on how audio features and strategies help create meaning in multimodal texts. The discussion is divided into three parts: audio features, audio strategies, and the study's research focus.

2.1 Audio Features

Sound in documentaries includes narration, dialogue, background music, sound effects, and ambient noise. It also involves features like loudness, pitch, timbre, and rhythm. These sounds do not just support the visuals—they help tell the story, build emotions, and express culture. Dialogue helps show realism and character. Changes in tone and rhythm make speech feel more authentic and help connect with the audience. Volume and pauses can change how urgent or intimate the speech feels. In multilingual settings, subtitles often work with dialogue to aid understanding. Narration links the visuals and audio. It provides background information and guides the viewer through the story. Pitch and tone affect how trustworthy or emotional the voice sounds. For example, a bright voice can build trust, while a softer tone can feel more personal. Rhythm and pauses often align with scene changes or key moments (Groves et al., 2025).

Music sets the mood and adds symbolic meaning. Groves et al. (2025) noted that fast rhythms and high notes create excitement, while slow and low music brings calm or seriousness. Bateman (2023) found that traditional music builds cultural resonance and connects with a wide audience. Music often follows the rhythm of the visuals to keep the flow smooth. Sound effects add realism and emotion. Loud, sharp sounds like glass breaking can create tension, while softer sounds build subtle feelings. Repeated patterns, such as ticking or drumming, help convey suspense or ritual. Ambient noise makes scenes feel real. Chion (2021) explained how natural sounds—like wind or crowds—immerse the viewer in the setting. Shifts between noise and silence help guide attention and emotion. Each sound feature plays a role. Loudness highlights key moments. Pitch conveys emotion. Timbre adds tone and texture. Rhythm controls pacing. These features often work with visuals and text, not independently. Many scholars also stress the need to balance sound with other modes. van Leeuwen (2021) warned against sound overload. Others, such as Groves et al. (2025) and Bateman (2023), emphasized that sound must match the visuals for coherence and impact.

2.2 Audio Strategies

Sound in documentaries is designed not only to support visuals but also to guide meaning, shape emotions, and reflect cultural values. Scholars widely agree that features such as pitch, loudness, timbre, and rhythm play an essential role in how audiences interpret auditory modes. van Leeuwen (1999, 2021) emphasizes that sound operates as a semiotic resource, creating symbolic and affective meaning much like images or text. Chion (2021) also stresses the importance of film sound in anchoring viewer perception, showing how silence, ambient noise, or sudden bursts of sound can direct attention and construct narrative rhythm.

Music is one of the most recognizable strategies. Groves et al. (2025) demonstrate that acoustic features of film scores—such as tempo and pitch height—evoke distinct emotional responses, from excitement to solemnity. Similarly, Kassabian (2001) highlight how background music frames cultural identity and emotional tone, while Bateman (2023) observes that traditional soundscapes in documentaries help establish cultural authenticity.

Narration remains another central strategy. Prosodic features like intonation, rhythm, and pauses determine whether narration sounds authoritative or intimate. Couper-Kuhlen (2021) shows that prosody in speech acts as dialogic action, shaping interaction and engagement. These findings

suggest that varying vocal delivery in narration not only conveys information but also builds trust and empathy across different audiences.

Ultimately, effective audio strategies rely on balance. Too much emphasis on one feature, such as timbre, risks overshadowing other semiotic resources (van Leeuwen, 2021). Instead, successful documentaries coordinate multiple auditory elements with the visual mode, ensuring coherence in storytelling and resonance with audiences across cultural contexts.

2.3 Research Focus

The first chapter listed the research questions. This section shows how this study builds on past work and explains its focus more clearly. Many researchers agree that sound in documentaries carries meaning. They have described its main features and strategies. However, most studies do not show how sound works with pictures and words at the same time, especially in Chinese festival documentaries.

This study focuses on how sound in Festive China helps tell stories, build emotions, and link with visuals and text. It looks at both how sounds are made (like pitch or rhythm) and how they are used (like mood or timing). The study treats sound as a key part of meaning—not just a background tool. By doing this, the study hopes to expand the way people look at meaning in multimodal texts. It gives special attention to sound in culturally rich documentaries. The next chapter explains how the study was done.

3. Research Methods

This study uses a qualitative method based on multimodal discourse analysis. It focuses on both how people use language and images to make meaning and how themes appear across different episodes. The goal is to explore how the English-language documentary *Festive China* uses sound, images, and words together to share culture, tell stories, and create emotional impact.

The analysis looks at how different types of meaning—such as what is shown, how it feels, and how it is arranged—come from the mix of spoken narration, background music, visuals, and ambient sounds. A thematic approach is also used to find common symbols and repeated design choices throughout the documentary.

To better understand how these kinds of texts can be used in learning, the study also asked students to write reflections. They answered an open question about any problems they had while analyzing the documentary. Their feedback helps show what makes it hard to understand and explain multimodal content.

3.1 Corpus of the Study

The main material for this study comes from the twelve-episode documentary *Festive China*. This documentary provides a rich and varied source for examining how Chinese festivals and traditions are presented through sound and image. Each episode centers on a specific holiday, such as the Spring Festival or the Dragon Boat Festival, and captures the unique customs, rituals, and seasonal atmosphere of that celebration. Together, the episodes create a broad picture of Chinese festive culture across different times of the year.

The documentary blends spoken narration, traditional and modern music, ambient soundscapes, visual scenes, and written language. These modes work closely together to explain cultural practices, guide emotional response, and build a deeper sense of connection between the audience and Chinese heritage.

Table 1 below lists the titles and lengths of all twelve episodes. It shows the full scope of the material used in the analysis.

Table 1. 12 Episodes in the Documentary *Festive China*

No.	Topics	Time (MM:SS)
Episode 1	The General Introduction of 24 Solar Terms in China	4:19
Episode 2	The History and Introduction of Solar Terms in Spring	3:55
Episode 3	The History and Customs of Clear Bright, Known as Qingming	3:30
Episode 4	The History and Introduction of Solar Terms in Summer	3:29
Episode 5	The History and Customs of Dragon Boat Festival	4:11
Episode 6	The History and Customs of Qixi Festival	4:06
Episode 7	The History and Introduction of Solar Terms in Autumn	3:52
Episode 8	The History and Customs of Mid-Autumn Festival	3:45
Episode 9	The History and Customs of the Double Ninth Festival	4:15
Episode 10	The History and Introduction of Solar Terms in Winter	3:52
Episode 11	The History and Customs of Spring Festival	4:13
Episode 12	The History and Customs of Lantern Festival	3:51
Total		57:12

3.2 Auditory Data

This study treats the audio in *Festive China* as an important meaning-making tool, just like the images and words. The sound design includes traditional music, background noise, and moments of silence. These sounds help build the mood, guide emotions, and support the story. For example, short pieces of music often appear when the visuals show rituals or seasonal scenes. Together, they create a strong and clear cultural message.

The analysis looks at how sound works in different ways. Some sounds set the tone. Others make the meaning stronger or help show where and when events happen. These audio choices make the film feel more real and help the story feel complete.

3.3 Analytical Framework

The study uses a method based on Systemic Functional Linguistics (SFL), developed by Halliday, and later applied to images and sound by Kress and van Leeuwen. This method focuses on three kinds of meaning: what is being shown (representational), how the viewer is involved (interpersonal), and how everything is put together (compositional).

SFL is particularly suitable for this dataset because it treats all semiotic resources—including language, image, and sound—as meaning-making systems that serve distinct but interrelated functions (Halliday, 1989). In the context of *Festive China*, festivals are not only narrated through words and visuals but also performed and reinforced through sound. Using SFL allows us to interpret how auditory choices contribute to ideational meaning (cultural content), interpersonal meaning (engagement with the viewer), and textual meaning (coherence across scenes).

For the sound part, the study also uses ideas from van Leeuwen about how to read meaning in audio. It looks at five features: how loud the sound is, its pitch, what it sounds like (timbre), how fast or slow it is (rhythm), and where it seems to come from (perspective). These features are not just technical. They also express feeling, create order, and show how close or far things feel.

Van Leeuwen's framework is especially appropriate here because it provides a systematic way to analyze sound as a semiotic mode, linking technical parameters (e.g., loudness, pitch) to social and cultural meanings (van Leeuwen, 1999, 2004). Unlike general film studies approaches that often treat sound as background or aesthetic effect, this framework conceptualizes audio as an integral resource for discourse, which is crucial for a documentary like *Festive China* that combines narration, traditional music, and ambient sounds to construct cultural identity.

By combining SFL and van Leeuwen's audio analysis, this study can capture both the broader functional orientation of multimodal discourse and the fine-grained features of sound design. This integrated approach makes it possible to explain how *Festive China* uses auditory resources not only to match visuals but also to carry cultural symbolism, shape audience engagement, and achieve coherence across episodes.

This approach helps explain how sounds and visuals work together. For instance, narration is studied by looking at tone and rhythm to show how it gives authority or friendliness. Music is studied to see how it supports culture and feelings. Background sounds like firecrackers or birds help the viewer feel like they are really there. This method shows how all the parts connect and support each other to tell a full cultural story.

3.4 Coding and Annotation Procedure

The study used a step-by-step process to look at how sound helps tell the story. First, each episode was divided into smaller parts based on topic, festival, or big changes in sound and image. Then, each part was checked for three types of sounds: narration, background music, and ambient noise. These sounds were chosen based on how often they appeared or how important they were to the story.

Each sound was then studied using five main features: loudness, pitch, timbre, rhythm, and the feeling it gives. Notes were written in a chart, with times and short descriptions of what was happening in the video. For example, a serious narration about ancestors might be paired with slow music and dark lighting, showing how the sounds and images work together.

Some parts of the narration were also written down to better study the tone and how it fits with the visuals. This careful, scene-by-scene method helped make sure each sound was studied in context, showing clearly how it adds to the film's message and meaning.

4. Results and Discussion

4.1 Typology and Features of Audio Modality

This section looks at how sound is used in the English documentary *Festive China*. It focuses on three main sound types: narration, background music, and ambient noise. The study explores how features like pitch, rhythm, timbre, and loudness help build cultural meaning. These sound elements work with images and words to make the story clearer, create emotion, and show cultural symbols. Together, they shape how viewers understand the film. The findings show that sound is not just background—it is a key part of how the documentary communicates meaning across cultures.

4.1.1 Narration

The narration in *Festive China* plays a central role in guiding the story and shaping cultural meaning. It uses a clear tone, steady pace, and simple but rich language to present information about Chinese festivals, while also adding emotional resonance to sustain audience engagement.

Table 2 presents selected examples of narration features across episodes. These examples illustrate how specific combinations of loudness, pitch, timbre, and rhythm work in different cultural contexts. The complete dataset is included in the Appendix A.

Table 2. Selected Examples of Narration Features in *Festive China*

Episode	Scene	Key Features	Function
1. The 24 Solar Terms	Explanation of origins	Medium loudness, steady rhythm, warm timbre	Ensures clarity and authority in cultural explanation
3. Clear and Bright	Tomb-sweeping rituals	Low pitch, solemn timbre, slow rhythm	Constructs reverence and ritual solemnity
11. Spring Festival	Fireworks celebration	Higher pitch, brighter tone, faster rhythm	Creates excitement and communal joy

From a systemic functional perspective, narration functions ideationally by conveying knowledge of rituals, seasonal changes, and traditional practices, interpersonally by positioning the narrator in relation to the audience, and textually by organizing the flow of the multimodal discourse. The medium loudness and steady rhythm in The 24 Solar Terms episode, for instance, establish an informative and authoritative stance, ensuring clarity while avoiding excessive intensity. In contrast, the lowered pitch and solemn timbre in Clear and Bright not only match the ritual scenes but also construct a discourse of reverence for ancestors. Similarly, the brighter tone and faster pace in Spring Festival foster an energetic, celebratory atmosphere that aligns with the visuals of fireworks and family gatherings. These variations demonstrate how narration moves beyond simple storytelling to articulate broader cultural meanings and values.

Van Leeuwen's framework further highlights how pitch, timbre, and rhythm serve as semiotic resources rather than technical parameters. The warm and intimate timbre in Mid-Autumn Festival creates a sense of familial closeness, functioning ideationally to foreground kinship and interpersonally to invite empathetic involvement. Meanwhile, rhythmic acceleration in Dragon Boat Festival does not merely support the excitement of boat races but textually synchronizes with music and ambient sounds to produce a multimodal cohesion of energy and collective pride. Through such patterns, narration contributes to constructing Chinese cultural identity as both historically grounded and emotionally accessible.

Overall, narration in Festive China cannot be reduced to a neutral voice-over. It operates as a multimodal semiotic resource that integrates with images, music, and ambient sound to produce layered cultural meanings. By flexibly adapting loudness, pitch, timbre, and rhythm to different festival contexts, the narration reinforces the documentary's dual role as both an informative account and a cultural performance aimed at international audiences.

4.1.2 Background Music

The background music in Festive China is meticulously designed to shape the emotional atmosphere and reinforce the cultural narratives of each episode. By modulating loudness, brightness, pitch, timbre, and rhythm, the music not only mirrors the mood of the visuals but also functions as a semiotic resource that enhances the documentary's storytelling.

Table 3 presents selected examples of background music features. These cases illustrate how the documentary aligns specific musical parameters with cultural contexts. The complete dataset is included in the Appendix A.

Table 3. Selected Examples of Background Features in *Festive China*

Episode	Scene	Key Features	Function
3. Clear and Bright	Tomb-sweeping rituals	Low pitch, soft timbre, slow rhythm	Creates solemnity and respect for ancestors
5. Dragon Boat Festival	Boat racing	High loudness, vibrant timbre, rapid rhythm	Produces excitement and collective energy
8. Mid-Autumn Festival	Moon-viewing	Medium loudness, gentle timbre, smooth rhythm	Evokes intimacy and familial harmony

From a systemic functional perspective, background music contributes to the ideational metafunction by representing cultural practices through sonic imagery, such as the solemn flute-and-piano ensemble accompanying ancestral rites in *Clear and Bright*. It also enacts the interpersonal metafunction, positioning the viewer emotionally through shifts in loudness and brightness—for example, the vibrant percussion in *Dragon Boat Festival* constructs an affective bond of excitement and pride. At the textual level, rhythm and timbre synchronize with narration and visuals, ensuring coherence and multimodal cohesion, as seen in the flowing strings of *Mid-Autumn Festival* that align with the visual calm of family gatherings.

Van Leeuwen's framework emphasizes that sound parameters such as pitch, timbre, and rhythm should be treated as semiotic resources rather than purely acoustic elements. For instance, the high-pitched celebratory gongs in *Spring Festival* not only signal joy but also index cultural meanings of renewal and prosperity. Similarly, the nostalgic warmth of traditional string instruments in *Autumn Harvest* constructs a discourse of continuity between past and present, reinforcing the cyclical rhythm of agrarian life. Through such patterned deployment, background music in *Festive China* transcends ornamental function and becomes integral to the cultural and emotional semiotics of the documentary. Overall, background music works in tandem with narration and ambient sound to immerse viewers in the multisensory world of Chinese festivals. By flexibly adapting sonic features to diverse contexts, it ensures that the auditory layer of the documentary is not simply supportive but actively generative of cultural meaning.

4.1.3 Ambient Sound

Ambient sound in *Festive China* enriches the auditory layer of the documentary by embedding viewers within authentic environments. It draws on natural sounds, daily activities, and festive noise to create a sense of presence that supports the documentary's cultural storytelling. Rather than serving as background filler, ambient sound functions as a semiotic resource that interacts with narration and music to produce coherence and immersion.

Table 4 presents selected cases of ambient sound. These examples illustrate how the documentary mobilizes sonic features to reflect cultural settings and evoke emotional resonance. The full dataset is provided in the Appendix A.

Table 4. Selected Examples of Ambient Sound Features in *Festive China*

Episode	Scene	Key Features	Function
1. 24 Solar Terms	Rustling leaves, gentle wind	Soft loudness, warm timbre	Evokes natural harmony and agrarian labor
5. Dragon Boat Festival	Splashing water, synchronized rowing	Moderate-to-loud, sharp timbre, fast rhythm	Conveys excitement and collective movement
11. Spring Festival	Firecrackers, distant cheers	Loud, bright timbre, rapid rhythm	Constructs celebratory energy and cultural festivity

From a systemic functional perspective, ambient sound fulfills multiple metafunctions. At the ideational level, it represents cultural and environmental processes: the rustling leaves in *The 24 Solar Terms* symbolize seasonal change and agricultural rhythms. At the interpersonal level, dynamic loudness positions the viewer emotionally, as the celebratory firecrackers in *Spring Festival* generate excitement and communal joy. At the textual level, rhythmic alignment with visuals—such as the synchronized rowing in *Dragon Boat Festival*—ensures temporal cohesion and multimodal fluency.

Van Leeuwen (2006) emphasizes that ambient sounds are not neutral but semiotically loaded. For example, firecrackers are not merely “loud noises” but culturally marked sounds that index renewal and auspiciousness in Chinese tradition. Similarly, the crisp crunching of snow in *Winter* (Episode 10) signals both environmental authenticity and symbolic purity. By carefully orchestrating these soundscapes, *Festive China* transforms environmental audio into carriers of cultural meaning.

Machin (2010) has argued that sound can create atmosphere as effectively as visual imagery. This is evident in *Festive China*, where cicadas and frogs in *Summer* embody vitality and abundance, while the subdued wind in *Double Ninth Festival* conveys solemn reflection. However, the documentary avoids the pitfall of sound overload: instead of overwhelming the visuals, ambient sound is balanced with narration and music, ensuring that cultural messages remain clear and focused. When compared with *Won't You Be My Neighbor?* (Neville, 2018), the function of ambient sound diverges. While the latter employs archival recordings to construct intimacy and biographical depth, *Festive China* deploys ambient sound to highlight collective traditions and cultural continuity. This contrast demonstrates that ambient sound is not universally functional but contextually adaptive, shaped by the documentary's communicative goals.

Overall, ambient sound in *Festive China* is a crucial meaning-making resource. By linking natural and cultural soundscapes to the multimodal narrative, it grounds the documentary in lived experience while reinforcing the symbolic weight of Chinese festivals.

4.2 Meaning Construction Through Auditory Features

The documentary *Festive China* uses sound in a smart and careful way. It mixes five basic audio features—loudness, brightness, pitch, timbre, and rhythm—to help tell each story and bring emotion into the scene. These sound choices work closely with images and words. Together, they create a rich cultural experience. This idea follows van Leeuwen's (2004) view that sound is not just background noise. It is a key part of how meaning is made. For example, changes in pitch and rhythm help express joy, calm, or respect during different festivals. Each sound feature is chosen to match the topic and feeling of the moment. This makes the audience feel more connected to the story.

4.2.1 Loudness

Festive China uses loudness in three main ways: through narration, background music, and ambient sound. Each one plays a special part in shaping how the audience listens and feels. The narrator speaks in a medium volume across all episodes. This choice helps keep the message clear and steady. It also avoids making the voice feel too strong or too soft. In Episode 1, *The 24 Solar Terms*, this even tone supports the teaching style. It helps viewers follow the historical and cultural points without stress.

In contrast, the music changes volume to match each scene's mood. For example, in *The Dragon Boat Festival*, the sound gets louder during the boat races. This makes the moment feel exciting and alive. In quieter scenes, like *Clear and Bright*, soft music supports the calm mood. It makes space for feelings of memory and respect. Ambient sound also uses volume to build the

world around each scene. Gentle sounds, like leaves moving in *The 24 Solar Terms*, bring peace. Sharp noises, like firecrackers in *Spring Festival*, bring energy. These changes help the sound match the visual mood.

In short, loudness helps balance emotion and clarity. Chion (2021) pointed out that changes in volume can draw people into the story and shape what they feel. Still, Bateman (2023) warned that too much change in volume might pull attention away from the other parts of the film, like visuals or voice. In *Festive China*, the volume choices stay controlled. They guide the audience without taking over. Beyond functional clarity, loudness also encodes cultural semiotics in *Festive China*. The restrained narration volume reflects Confucian ideals of balance and moderation, while sudden loud bursts—such as firecrackers—symbolize renewal and collective vitality in ritual practice. In this sense, volume is not only a technical adjustment but a cultural marker that links sonic intensity to shared symbolic meaning.

4.2.2 *Brightness*

Brightness, or the sharpness and clarity of sound, changes with the tone of each episode in *Festive China*. It helps set the mood and gives each scene a clear emotional color. The narrator's voice becomes brighter in happy moments. In *The Lantern Festival*, for example, the lively brightness of the voice matches the glowing lanterns and joyful scenes. In quieter episodes, like *Clear and Bright*, the voice softens. The dimmer tone fits the quiet mood of remembering ancestors.

The music also uses brightness to match feelings. In *The Lantern Festival*, the melodies are clear and full of light, adding to the fun and celebration. But in more serious scenes, like *The Double Ninth Festival*, the music becomes soft and low in brightness. This change adds a feeling of memory and thoughtfulness. Natural sound works the same way. High and sharp noises, such as cicadas in *Summer*, bring out the warmth and life of the season. In *Winter*, muffled footsteps on snow feel quiet and deep, matching the cold and calm setting. These choices make each moment feel more real. Chion (2021) said that sound brightness can guide how we feel about a scene. But Bateman (2023) reminded us that if brightness is too strong, it can hide smaller sounds and make the story feel less balanced. In *Festive China*, brightness is used with care. It helps the audience connect with both the mood and the meaning.

More importantly, the modulation of brightness in *Festive China* demonstrates how acoustic clarity is mobilized as a semiotic code for moral and aesthetic values. Brighter timbres in festive settings symbolically align with the Chinese semiotic association between luminosity and prosperity, while dimmer sounds in ritual contexts embody reverence and remembrance. This culturally grounded interpretation extends beyond technical description to reveal brightness as an auditory metaphor for the seasonal cycle of vitality and decline, a recurring theme in Chinese cultural discourse.

4.2.3 *Pitch*

Pitch helps build the feeling of each scene. High or low sounds guide the viewer's emotions and focus. The narrator uses lower pitch when talking about serious topics. For example, in *Clear and Bright*, the voice becomes deeper. This fits the calm and respectful mood. In contrast, higher pitch is used in joyful episodes like *Spring Festival*. The lighter voice matches the happiness of family gatherings. Music does the same thing. In *Spring*, high-pitched tunes sound fresh and full of life. They support the ideas of growth and change. Lower-pitched music in *Clear and Bright* brings a sense of tradition and honor.

Ambient sounds also show pitch changes. Bird songs in Spring are high and cheerful. They make the season feel new and alive. On the other hand, in *The Double Ninth Festival*, low sounds like mountain wind create a quiet, thoughtful mood. Pitch helps the viewer feel the emotional shift between scenes. Chion (2021) believed that pitch is a strong tool for building deep stories through sound. Still, Van Leeuwen (2021) warned that using pitch alone can make stories feel too simple. In *Festive China*, pitch is just one part of a bigger sound plan. It works together with other features to keep the story balanced and rich.

The semiotic deployment of pitch in *Festive China* reveals a deeper symbolic stratification. High-pitched sounds are not only markers of joy but also indexical of renewal and youth in Chinese seasonal culture, whereas low-pitched tones resonate with gravity, ancestry, and historical depth. In this way, pitch functions as a cultural signifier rather than merely an acoustic variation. The documentary's careful alternation between registers underscores how tonal height metaphorically encodes cosmological dualities such as vitality versus solemnity, aligning sonic practice with broader cultural narratives of continuity and change.

4.2.4 Timbre

Timbre is the color of the sound. It tells us how a sound feels, whether warm, sharp, smooth, or rough. In *Festive China*, timbre changes to match each story. The narrator's voice shifts across episodes. In *Mid-Autumn Festival*, the timbre is soft and warm. It brings out the closeness of family. In *Qixi Festival*, the voice becomes clearer and more romantic. This supports the theme of love and longing.

Music timbre also plays a role. In *Autumn*, the music has a deep, soft tone. It feels full of memory and thanks. In *The Lantern Festival*, the sound is bright and shining, helping the joyful feeling grow stronger. Ambient sounds add to this. In *Autumn*, the rustling of rice fields has an earthy, rich tone. It connects the viewer to nature and farming traditions. In *Spring Festival*, the crackle of firecrackers is loud and sharp. It brings excitement and life to the scene.

Timbre helps establish links between emotion, culture, and setting. Research has shown that timbre can shape listeners' emotional responses to sound (Eerola etc., 2012). But van Leeuwen (2021) reminded that timbre should not take over. If we focus too much on tone, we might miss other important parts like rhythm or pitch. *Festive China* avoids this by using timbre with balance and care.

Beyond its descriptive variability, timbre in *Festive China* encodes culturally salient textures of life. The use of warm, earthy tonalities during harvest scenes resonates with agrarian values of rootedness and community, while bright, metallic timbres in celebratory rituals evoke the sonic aesthetics of spectacle and festivity in Chinese culture. Thus, timbre operates as a cultural index that fuses emotion with social practice, reinforcing the symbolic alignment between natural sound textures and collective cultural identity.

4.2.5 Rhythm

Rhythm controls how fast or slow the sound feels. It helps guide the pace of the story in *Festive China*. The narrator adjusts rhythm to match the topic. In *The Double Ninth Festival*, the rhythm is slow and calm. It fits the mood of honoring family and tradition. In *The Dragon Boat Festival*, the rhythm is quick and lively. It matches the fast rowing and excitement.

Music rhythm changes with the feeling of each scene. Fast beats in *Summer* add energy to harvest celebrations. In contrast, slow, flowing music in *Mid-Autumn Festival* supports the peaceful feeling of family dinners and moon watching. Even natural sound follows this idea. In *Qixi Festival*,

water flows gently, helping the scene feel romantic and still. In The Dragon Boat Festival, the beating of paddles is sharp and regular. This builds power and rhythm into the scene.

Rhythm keeps the story moving and helps show how each moment should feel. Chion (2021) said that rhythm holds the viewer's focus and brings images to life. But van Leeuwen (2021) warned that too much focus on rhythm might hide small details in the sound. Festive China finds a good balance. It lets rhythm lead the pace but still leaves space for other sounds to matter.

At a deeper level, rhythm in Festive China functions as a semiotic representation of temporal order in Chinese culture. The alternation between fast and slow beats mirrors the cyclical rhythm of agricultural labor and seasonal ritual, embedding the documentary's pacing within broader cosmological patterns of time. This suggests that rhythm is not only a narrative device but also an auditory embodiment of cultural temporality, aligning sound with the ritual calendar and the symbolic life cycle of the community.

The analysis of Festive China demonstrates that sound operates as a central semiotic mode rather than mere background. By orchestrating loudness, brightness, pitch, timbre, and rhythm, the documentary constructs emotional resonance and cultural depth. Loudness and brightness frame shifts between festive excitement and reflective calm, while pitch and timbre infuse symbolic meaning aligned with tradition and setting. Rhythm integrates these features, sustaining narrative flow and guiding audience attention. Together, these auditory choices extend beyond surface-level scene matching, revealing how sound mediates cultural expression and enhances viewer immersion. This supports van Leeuwen's notion of sound as meaning-making and illustrates how multimodal strategies shape collective memory and identity in contemporary Chinese documentary practice.

4.3 Modal Interaction and Multimodal Synergy

In Festive China, the combination of narration, background music, and ambient sound builds a rich and layered listening experience. These audio elements work closely together to bring out the cultural meaning and emotional feeling of each festival. At the same time, they connect smoothly with the images and spoken content, helping viewers fully experience the depth and beauty of Chinese traditions.

Narration as the Anchor of Cultural Context. Narration serves as the foundation of the auditory experience, providing clear and accessible cultural and historical insights. Controlled loudness, steady rhythm, and adaptive pitch ensure that the narration is both informative and emotionally engaging. For instance, the medium loudness and measured rhythm in episodes like The 24 Solar Terms maintain clarity and accessibility, complementing the educational tone. Narration gains additional depth through its interaction with background music and ambient sound. In reflective episodes like Clear and Bright, the solemn tone of the narration pairs with subdued music and soft ambient sounds, such as rustling leaves, to create a layered auditory atmosphere. Conversely, celebratory episodes like Spring Festival feature energetic narration, vibrant music, and dynamic ambient sounds, such as firecrackers, which amplify the festive mood.

Background Music as an Emotional Amplifier. Background music is carefully tailored to reflect the thematic and emotional tone of each festival. Its variations in loudness, pitch, timbre, and rhythm enrich the narrative and evoke emotional responses. For example, high-pitched, lively melodies in Spring emphasize themes of renewal and vitality, while resonant and nostalgic tones in Mid-Autumn Festival underscore family unity and gratitude. The interplay between music and narration creates a synergistic effect. In The Dragon Boat Festival, rhythmic beats mirror the energetic rowing competitions described in the narration, immersing viewers in the festival's vibrancy. Similarly, in The Double Ninth Festival, the gentle rhythm and soft timbre of the music

complement the reflective tone of the narration, fostering an atmosphere of reverence and introspection.

Ambient Sound as a Connector to the Environment. Ambient sound provides a sensory link to the cultural and natural environments depicted in the documentary. Sounds like chirping cicadas in Summer, crackling firecrackers in Spring Festival, and flowing water in Qixi Festival enhance authenticity and immersion. These auditory elements situate viewers within specific scenes while reinforcing the emotional tone set by the narration and background music. For instance, in *Clear and Bright*, the soft rustling of leaves and muted sounds of tomb-sweeping rituals complement the solemn narration and subdued music, creating a reflective and respectful auditory landscape. In *Lantern Festival*, the vibrant sounds of bustling streets and glowing lanterns amplify the celebratory tone of the narration and bright melodies, enveloping viewers in the festival's joyous spirit.

Dynamic Interplay Across Modalities. The dynamic interplay between narration, background music, and ambient sound exemplifies the synergy within Festive China's audio modalities. Each element is purposefully designed to enhance the cultural narrative while remaining distinct and complementary. This interplay ensures a cohesive, multilayered auditory experience that captures both the grandeur and intimacy of Chinese festivals. In *Spring Festival*, the energetic tone of the narration is elevated by fast-paced music and the vibrant sounds of fireworks, creating a lively and celebratory auditory landscape. In contrast, *The Double Ninth Festival* features a reflective tone supported by soft music rhythms and the calm ambient sound of autumn winds, emphasizing themes of longevity and reverence for elders.

Cultural and Emotional Resonance Through Synergy. The integration of audio modalities enhances the cultural and emotional resonance of the documentary. By weaving narration, background music, and ambient sound into a cohesive auditory narrative, Festive China ensures that audiences not only hear the story but also feel its emotional and cultural depth. For example, the harmonious blending of narration, music, and soundscapes during *Mid-Autumn Festival* evokes a profound sense of family unity and nostalgia. Similarly, the vibrant interplay of these modalities during *The Dragon Boat Festival* captures the festival's energetic spirit and communal pride. This synergy bridges cultural boundaries, inviting a global audience to connect with the richness of Chinese traditions.

The analysis shows that the use of different audio elements in Festive China works well to build both cultural meaning and emotional depth. These elements—narration, background music, and ambient sound—are carefully combined to support the story of each festival. This matches what Chion (2021) pointed out: when sound elements are used together, they help create a rich and engaging listening experience.

At the same time, not all scholars fully agree. van Leeuwen (2021) warned that focusing too much on sound combinations might make it harder to notice the role of each part. If one element becomes too strong, it could weaken the others. Even so, Festive China shows how to keep things in balance. The audio is blended in a way that supports the visuals and the message. By putting the different sounds together in a thoughtful way, the documentary brings out both the emotional feeling and the cultural meaning of each scene.

Beyond descriptive synergy, the interaction of modalities also points to broader semiotic principles. Narration anchors ideational meaning, music intensifies interpersonal affect, and ambient sound strengthens contextual embedding. Their orchestration demonstrates how auditory modalities can map onto Halliday's metafunctions, suggesting that Festive China's design is not only aesthetic but also structurally systemic. This theoretical framing elevates the analysis from cataloguing sound effects to identifying functional roles in multimodal discourse.

4.4 Auditory Balance and Aesthetic Constraints

In Festive China, the sound design focuses on creating a calm and balanced listening experience. The documentary carefully avoids strong emotional extremes, following traditional Chinese ideas of moderation and harmony. This approach helps each scene match the mood of the festival—whether joyful or reflective—without overwhelming the viewer.

One way the film achieves this is by avoiding sounds that are too loud or too emotional. Instead of using intense music or dramatic effects, it often chooses peaceful and steady sounds. For example, in the Spring Festival episode, the music sounds bright and cheerful, but it stays controlled. The rhythm is lively but not too fast. This creates a happy feeling while keeping the mood gentle and comfortable.

This sense of balance also appears in how loudness and rhythm are handled. In the Mid-Autumn Festival scene, soft background music plays behind quiet family conversations. The music stays in the background, so viewers can focus on the voices and natural sounds. The slower rhythm fits the mood of reunion and reflection, highlighting the close ties between family members.

On the other hand, during the Dragon Boat Festival, the volume increases to match the excitement of the race. The sound of paddling and cheering grows stronger, and the rhythm speeds up. Even so, the sound never becomes too loud or chaotic. It stays in line with the story and keeps the energy under control. Through this careful sound design, Festive China builds an emotional bond with the audience while staying true to the values of balance and restraint.

This strategy reflects more than aesthetic preference; it embodies a cultural semiotic principle of harmony. Loudness and rhythm are not only technical adjustments but markers of symbolic restraint, aligning with Confucian ideals of moderation. By contrasting controlled intensity in festive moments with subdued tones in reflective ones, Festive China stages an “auditory ethics” where sound carries moral as well as affective weight. Such an interpretation highlights that the documentary’s auditory design is simultaneously cultural performance and normative discourse, rather than a neutral aesthetic choice.

5. Conclusions

5.1 Summary and Interpretation of Key Findings

This study explored how sound helps shape cultural meaning in the English-language documentary Festive China. Using Systemic Functional Linguistics and multimodal discourse analysis as the main framework, it focused on three kinds of audio—narration, background music, and ambient sound. Each was examined through five key features: loudness, pitch, timbre, rhythm, and brightness. The results show that these sounds do more than support the visuals. They play a central role in creating meaning on three levels: ideas, relationships, and composition. For example, narration not only explains but also builds an emotional link with viewers. Background music adds mood and supports the story's flow. Ambient sounds make scenes feel more real and help express cultural values.

These findings suggest that auditory design in cultural documentaries functions not merely as accompaniment but as a meaning-making resource in its own right, mapping onto Halliday’s three metafunctions and thereby expanding the theoretical scope of multimodal discourse analysis.

5.2 Contributions and Implications

This research brings more attention to audio in multimodal discourse analysis. It shows that sound can carry meaning just like images or text. While many past studies focus on what we see

and read, this study highlights what we hear. Through detailed analysis, it shows how documentaries use sound to express culture and guide how audiences feel and understand. In education, these insights can help students build stronger skills in multimodal literacy. Teachers can use sound-focused tasks to help learners think more deeply about media, especially when studying films or programs with rich cultural content.

Beyond pedagogy, the findings also have implications for media production and cultural communication. Documentary makers can deliberately design soundscapes to convey values of harmony, reverence, or festivity, while cultural institutions may employ similar strategies to enhance cross-cultural understanding and soft power projection.

5.3 Limitations and Directions for Future Research

Although this study offers a clear picture of how sound works in Festive China, it mainly relies on qualitative methods. It does not include detailed, time-based analysis using tools like ELAN. Future research could use such tools to track when and how different sounds appear, how often they occur, and how they combine. Also, looking at documentaries from other cultures or languages could help us see how sound is used in different ways. Lastly, more work is needed on how viewers understand and react to these sounds. Studying audience feedback could show how sound shapes not just meaning, but also emotional responses.

Future research may also benefit from combining reception studies with computational methods to link sound design features with measurable audience responses, thereby bridging the gap between semiotic analysis and empirical validation. Together, these findings show that sound plays a key part in building cultural meaning. To fully understand multimodal discourse, future studies should treat audio as just as important as what we see and read. By foregrounding sound as a semiotic mode, this study not only enriches multimodal theory but also repositions auditory meaning-making as a crucial site for understanding how culture is experienced and transmitted across media.

References

- Bateman, J. A. (2023). *Multimodal discourse analysis: Theories and methods*. Cambridge University Press.
- Chion, M. (2021). *Audio-vision: Sound on screen* (Updated ed.). Columbia University Press.
- Couper-Kuhlen, E. (2021). Prosody as dialogic action: Evidence from speech. *Frontiers in Psychology, 12*, 640741. <https://doi.org/10.3389/fpsyg.2021.640741>
- Dong, M., & Yuan, X. L. (2021). On constructing a multimodal aesthetic critical discourse analysis framework. *Foreign Language Education, 42*(1), 77–82. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2021.01.014>
- Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception, 30*(1), 49–70. <https://doi.org/10.1525/mp.2012.30.1.49>
- Feng, W. D. (2017). Metonymy and visual representation: Towards a social semiotic framework of visual metonymy. *Visual Communication, 16*(4), 441–466. <https://doi.org/10.1177/1470357217717142>
- Groves, K., Farbood, M. M., Carone, B., Ripollés, P., & Zuanazzi, A. (2025). Acoustic features of instrumental movie soundtracks elicit distinct and mostly non-overlapping extra-musical meanings in the mind of the listener. *Scientific Reports, 15*, Article 2327. <https://doi.org/10.1038/s41598-025-86089-6>

- Halliday, M. A. K., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective* (2nd ed.). Oxford University Press.
- Jackson, B. (2022). Sound anchors: A cognitive and multimodal approach to sound and narrative structure in film. *Projections*, 16(3), 61–82. <https://doi.org/10.3167/proj.2022.160303>
- Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. Routledge.
- Kassabian, A. (2001). *Hearing film: Tracking identifications in contemporary Hollywood film music*. Routledge. <https://doi.org/10.4324/9780203427460>
- Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge. <https://doi.org/10.4324/9781003099857>
- Li, M., Yasin, M. A. I., Hashim, N. B., & Ang, L. H. (2020). A historical review of the Chinese documentary programs. *Journal of Critical Reviews*, 7(2), 723–732. <https://doi.org/10.31838/jcr.07.02.133>
- Machin, D. (2010). *Analysing popular music: Image, sound, text*. SAGE. <https://doi.org/10.4135/9781446280027>
- Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies*, 10(4), 347–355. <https://doi.org/10.1080/17405904.2013.813770>
- Martinec, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 7(3), 337–371. <https://doi.org/10.1177/1470357205055928>
- Nai, R., & Hassan, H. (2022). A multimodal discourse analysis of BBC documentary *Chinese New Year: The Biggest Celebration on Earth*. *Hong Kong Journal of Social Sciences*, (60), 402–411. <https://doi.org/10.55463/hkjs.issn.1021-3619.60.40>
- Nan, N. (2024). Recording China in sound: Audio-visual landscape and cultural expression in “China in the Sound.” *Education Reform and Development*, 6(3), 263–268. <https://doi.org/10.26689/erd.v6i3.6586>
- Neville, M. (Director). (2018). *Won't you be my neighbor?* [Film]. Focus Features.
- Pan, Y. Y., & Zheng, Z. H. (2018). On the criticality of multimodal discourse analysis. *Journal of Tianjin Foreign Studies University*, 25(2), 78–87.
- Tian, H. L., & Pan, Y. Y. (2018). From meaning to intention: New development of multimodal discourse analysis to multimodal critical discourse analysis. *Shandong Foreign Language Teaching*, 39(1), 23–33. <https://doi.org/10.16482/j.sdwy37-1026.2018-01-003>
- van Leeuwen, T. (1999). *Speech, music, sound*. Palgrave Macmillan.
- van Leeuwen, T. (2004). *Introducing social semiotics*. Routledge.
- van Leeuwen, T. (2021). *Multimodality and identity*. Routledge. <https://doi.org/10.4324/9781003186625>
- Wang, R. (2023). Multimodal analysis of image-text relations in the translation of news cartoons in China Daily. *Frontiers in Humanities and Social Sciences*, 3(8), 27–37. <https://doi.org/10.54691/fhss.v3i8.5534>
- Wei, X. (2024). Multimodal metaphors in a Sino-British co-produced documentary. *Humanities and Social Sciences Communications*, 11, Article 286. <https://doi.org/10.1057/s41599-024-02739-6>
- Zhao, J. (2020). Interpreting the cultural symbols in Chinese documentaries: Taking *A Bite of China* and *Beautiful China* as examples. *Frontiers in Educational Research*, 3(11), 41–46. <https://doi.org/10.25236/FER.2020.031107>

Appendix A

Table A1. The Features of Narration

Episode	Scene	Loudness	Brightness	Pitch	Timbre	Rhythm
1	Explaining the origin of solar terms and their connection to farming and festivals	Medium	Neutral	Medium	Warm and informative	Steady
	Depicting Winter Solstice rituals at the Temple of Heaven, featuring ceremonial activities	Low	Subdued	Low	Solemn and resonant	Slow
2	Describing the 'biting spring' tradition, where people eat fresh seasonal foods to celebrate	Medium	Bright	Medium-high	Cheerful and lively	Flowing
	Tea leaf picking in lush countryside hills, accompanied by the chatter and laughter of workers	Medium-low	Warm	Medium	Soft and dynamic	Moderate
3t	Family members performing tomb-sweeping rituals with offerings and prayers for ancestors	Medium-low	Neutral	Low	Reverent and clear	Slow
	Families enjoying spring outings, strolling through blooming landscapes in mild sunshine	Medium	Bright	Medium	Light and engaging	Gentle and flowing
4	Farmers harvesting ripe crops under the hot summer sun, showcasing bustling rural activity	Medium-high	Bright	High	Energetic and rich	Fast-paced
	Enjoying watermelons during summer gatherings, with people savoring the refreshing fruit	Medium	Warm	Medium-high	Light and soft	Flowing
5	Energetic dragon boat races on the river, with rhythmic paddling and cheering crowds	Medium-high	Bright	High	Vibrant and rhythmic	Pulsating
	Families preparing zongzi together, wrapping rice and fillings in a cozy and festive atmosphere	Medium	Neutral	Medium	Warm and intimate	Steady
6	Narrating the romantic legend of the Cowherd and Weaver Girl, focusing on their longing	Medium-low	Warm	Medium	Tender and reflective	Gentle and flowing
	Couples exchanging gifts and registering marriages as part of modern Qixi celebrations	Medium	Bright	Medium-high	Cheerful and lively	Moderate
7	Golden rice terraces shining in the autumn sun, with farmers harvesting amidst breathtaking scenery	Medium	Bright	Medium	Resonant and vibrant	Rhythmic and steady
	Traditional dances and ritual activities during harvest festivals in the Longji terraces	Medium-high	Warm	Medium-high	Lively and energetic	Moderate to fast

8	Families gathering under the full moon, sharing stories and snacks in an intimate setting	Medium-low	Warm	Medium-low	Intimate and serene	Slow and flowing
	Preparing and enjoying mooncakes, with detailed visuals of the process and shared enjoyment	Medium	Bright	Medium	Soft and warm	Gentle and steady
9	People climbing mountains for blessings, surrounded by vibrant autumn foliage	Medium	Neutral	Medium	Reflective and calm	Steady
	Admiring blooming chrysanthemums and sipping chrysanthemum wine as part of the celebration	Medium-low	Subdued	Medium-low	Soft and clear	Gentle and steady
10	Scenes of ice fishing on Chagan Lake, with tools breaking through frozen surfaces	Medium	Bright	Medium-low	Crisp and resonant	Moderate and flowing
	Families preparing preserved meats during Minor Snow, showing traditional culinary techniques	Medium-low	Warm	Medium-low	Subdued and rich	Slow and steady
11	Fireworks lighting up the sky on New Year's Eve, accompanied by laughter and celebrations	Medium-high	Bright	Medium-high	Vibrant and festive	Pulsating
	Adults and children exchanging red envelopes, symbolizing blessings and familial bonds	Medium	Neutral	Medium	Warm and cheerful	Gentle and rhythmic
12	Crowds admiring illuminated lanterns of various designs during nighttime festivities	Medium	Bright	Medium	Luminous and soft	Smooth and flowing
	Families sharing tangyuan, with visuals of steaming bowls and warm conversations	Medium-low	Warm	Medium-low	Warm and intimate	Gentle and rhythmic

Table A2. The Features of Background Music

Episode	Scene	Loudness	Brightness	Pitch	Timbre	Rhythm
1	Gentle strings accompany the narration describing the origins of the 24 Solar Terms.	Medium	Soft and warm	Low to medium	Smooth and mellow	Steady and flowing
2	Uplifting orchestral music highlights the renewal of life during springtime tea-picking scenes.	Medium to high	Bright and vibrant	Medium to high	Warm and clear	Light and rhythmic
3	Solemn piano and flute music during the ancestral tomb-sweeping scenes.	Low to medium	Soft and reflective	Low	Resonant and gentle	Slow and deliberate
4	Cheerful folk instruments like bamboo flutes during summer harvest and watermelon eating scenes.	Medium to high	Bright and cheerful	Medium to high	Lively and textured	Upbeat and dynamic

5	Energetic drums and fast-paced strings during dragon boat racing scenes.	High	Sharp and vibrant	Medium to high	Powerful and resonant	Rapid and driving
6	Romantic strings and harp melodies during scenes narrating the legend of the Cowherd and Weaver Girl.	Medium	Bright and dreamy	Medium to high	Delicate and lyrical	Gentle and flowing
7	Warm, earthy tones of traditional instruments during autumn harvest celebrations.	Medium	Soft and nostalgic	Low to medium	Rich and grounded	Calm and steady
8	Melodic flute and soft strings set a tranquil mood during moon-viewing scenes.	Medium	Gentle and luminous	Medium	Clear and soothing	Smooth and serene
9	Traditional wind and string instruments evoke reflection during scenes of mountain climbing.	Low to medium	Soft and balanced	Low to medium	Echoing and contemplative	Slow and steady
10	Deep string notes and chimes create a serene atmosphere during snowy landscapes.	Low to medium	Gentle and cold	Low	Hollow and soft	Slow and atmospheric
11	Vibrant drums and celebratory gongs during scenes of firework displays and family reunions.	High	Bright and energetic	Medium to high	Resonant and festive	Upbeat and celebratory
12	Luminous strings and tinkling bells during the scenes of glowing lantern exhibitions.	Medium	Bright and magical	Medium	Sparkling and clear	Flowing and light

Table A3. The Features of Ambient sound

Episode	Scene	Loudness	Brightness	Pitch	Timbre	Rhythm
1	Sounds of rustling leaves and gentle wind accompanying a scene of farmers working in the fields.	Soft	Warm	Mid-low	Natural and airy	Slow and tranquil
2	Tea-picking scene with faint bird songs and distant chatter echoing in the hills.	Soft to moderate	Bright and clear	Mid-high	Crisp and melodic	Flowing and relaxed
3	The sound of sweeping tombs with rustling leaves and distant temple bells.	Moderate	Dim and reflective	Low-mid	Resonant and solemn	Steady and reflective
4	Frogs croaking and cicadas chirping in a serene countryside evening.	Moderate	Bright and sharp	High	Crisp and buzzing	Rhythmic and lively
5	Splashing of water and synchronized rowing sounds during a dragon boat race.	Moderate to loud	Bright and vibrant	Mid-high	Sharp and energetic	Fast and dynamic
6	Gentle flowing water and soft whispers in a romantic garden scene.	Soft	Warm and soothing	Low-mid	Delicate and smooth	Slow and flowing
7	Wind blowing through golden rice fields with faint harvest chatter in the background.	Soft to moderate	Warm and earthy	Low-mid	Rustic and natural	Gentle and steady
8	Family reunion with faint background murmurs and occasional laughter under a full moon.	Moderate	Warm and intimate	Mid	Smooth and mellow	Relaxed and rhythmic

9	Rustling chrysanthemums and mountain wind during a climbing scene.	Moderate	Clear and crisp	Mid-high	Breezy and natural	Flowing and steady
10	Crunching snow and muffled footsteps in a frozen landscape.	Soft	Dim and muted	Low-mid	Crisp and cold	Slow and deliberate
11	Firecrackers and distant cheers marking the New Year celebrations.	Loud	Bright and vivid	Mid-high	Crackling and dynamic	Fast and energetic
12	Lantern-lit streets filled with murmurs, footsteps, and occasional laughter.	Moderate	Bright and lively	Mid-high	Smooth and festive	Rhythmic and steady