

From Seeing to Perceiving: The Audiovisual Construction and Embodied Experience of Rural Space in Documentaries

Qu Lei¹, Teo Miaw Lee¹ & Candida Jau Emang¹

¹ Faculty of Applied and Creative Arts, Universiti Malaysia Sarawak, 94300, Kota, Samarahan, Sarawak, Malaysia

Correspondence: Qu Lei, Faculty of Applied and Creative Arts, Universiti Malaysia Sarawak, 94300, Kota, Samarahan, Sarawak, Malaysia.

doi: 10.63593/SSSH.2709-7862.2025.11.001

Abstract

This paper discusses some recent Chinese documentaries with rural themes. These documentaries use audiovisual codes to help audiences move from “seeing” spaces physically to “sensing” multidimensional spaces, triggering embodied experiences. Owing to a general tendency to focus on “what kind of space is depicted” in documentaries, scholars have systematically neglected the basic link of “how they trigger audiences’ sensory experiences”. Taking some recent documentaries including *Remembering Nostalgia* and *Beautiful Countryside* as examples, this paper intends to argue that there is a central argument in these documentaries: That is, documentaries “guide” viewers visually and “envelop” them aurally, making them an attractive textual body. That is, first, visually, mobile shots “guide” audiences to experience space as if they were “walking”, close-ups “stimulate” audiences to feel emotions, and changes in lighting effect create spatial atmosphere, which turn viewers from merely “seers” of spaces into active “explorers” of spaces. Second, aurally, a carefully designed soundscape and affective voice-over create a surrounding acoustic space that “envelops” audiences. In sum, the entire audiovisual strategy not only depicts images of countryside, but also effectively “mobilizes” viewers’ bodily memories and affective feelings to trigger embodied experiences, that is, “being there”. By exploring the sensory aesthetics and experiential construction in documentaries, this paper reveals how documentaries can help audiences leap from “seeing” spaces to “sensing”.

Keywords: audiovisual construction, embodied experience, rural space, sensory aesthetics, Chinese documentaries

1. Introduction

1.1 Research Questions

This paper intends to explore in a systematic way how contemporary Chinese rural documentaries, through the creation of new audiovisual codes, deeply transform the representation of rural space into an emotional mobilization. The research question of this paper is: How do documentary creators use certain kinds of audiovisual codes to establish emotional channels between viewers and rural space so that viewers go beyond a mere visual consumption of rural space and can produce embodied spatial experiences and emotions?

The above-mentioned research question can be further concretely divided into three interrelated dimensions: First, in terms of visual construction, how do documentaries use camera language (including the organic combination of shot type, designed camera movement trajectory and style treatment of lighting) to transcend the planar limitation of frame and produce an immersive visual environment to guide viewers to have a “being-there” sensory experience? Second, in terms of auditory construction, how do documentaries use designed soundscapes (including the spatial reproduction of ambient sound, emotional guidance of voice-over

narration and the rhythmic control of sound elements) to produce more three-dimensional space and then evoke viewers' emotional memories and produce a holistic perception of rural space? Third, in terms of the integrated audiovisual level, how do visual and auditory elements use construct a kind of "sensible space" which is resonant with the discourse of rural development in the new era and also with the audience?

Those questions are not only related to the analysis of documentary techniques but also related to some theoretical issues concerning media communication, emotional mobilization and cultural construction.

1.2 Literature Review

As the deep media integration is gradually realized, how audiovisual media represent rural space has attracted increasing attention from scholars. Existing studies promote the academic discourse on this issue from three dimensions.

First, existing studies focus on the representation of rural space and its narrative turn. Wang Yuwei and Yang Shan (2024) systematically reviewed the narrative image of rural areas in the media, and found that its image has undergone a significant transformation from a poetic "utopian" imagination to a critical "pre-modern" symbol, and then to the "livable, workable, beautiful, and harmonious" image of the new era. The narrative turn of the image of rural areas reveals the deep connection between media representation and social ideology. From the perspective of communication studies, Zhang Huiyu (2024) found that the audiovisual technology has brought new visual impact and emotional experience to the cultural communication of rural areas, and enabled the rural areas to shift from a "disadvantaged" position to a representative of "livable, harmonious, and valuable" spaces in the digital age.

In terms of audiovisual technique, scholars have conducted relatively in-depth exploration. Wang Xinying and You Yaowang (2024) systematically analyzed the audiovisual presentation of documentaries on the revitalization of rural areas, and explained the basic function of technical elements such as shot selection, lighting application, and sound design in spatial construction in detail. Zhao Li (2024) expanded the scope of analysis to more visual media, and analyzed the overall use of camera language, editing rhythm, and sound design in the discourse of rural areas, while emphasizing that the audiovisual media can reconstruct the cultural value of rural areas.

The above-mentioned theoretical breakthroughs in technique have laid a solid foundation for the analysis of documentary audiovisual grammar. More importantly, the theoretical model of multidimensional space has been introduced. Starting from the spatial theory of Lefebvre and Hafique, Wang Yuwei and Yang Shan (2024) constructed a tripartite analytical model of material space, spiritual space, and social space. This theoretical model breaks through the traditional understanding of rural areas as a single geographical space, redefines the rural areas as a complex field filled with cultural meaning and social relations, and provides a powerful theoretical model for the in-depth interpretation of rural areas.

Although they have made these great achievements, current research finds a theoretical blind spot: most of the existing studies remain at the descriptive level of what documentaries construct and which techniques they use. Very few studies focus on the deep level of how form and content jointly influence the audience's sensory experience. Existing research is very good at analyzing the "representation" of space, but they seldom explore the "experience" of space; they are very good at explaining the symbolic meaning of audiovisual signs, but they do not clearly reveal how audiovisual signs mobilize the audience's bodily perception to make them feel a "sense of immersion" or an embodied experience. This necessary transition from "representation" to "experience" is what this article intends to promote.

Based on the existing research on representation of space and audiovisual technique, this article will first present the theoretical perspective of "embodied experience". Then, based on careful analysis of the accurate audiovisual syntax of contemporary rural documentaries, this article will explore how they transform macro-level discussions of countryside into emotional experience that can be felt bodily by audience. This is a research path that transcends the usual question of "what kind of countryside does documentary show?", and explores further "how does documentary make the audience feel that kind of countryside with their whole being?". This is what this article can contribute to current documentary studies on sensory aesthetics and mechanism of experience construction.

1.3 Research Methods

This study adopts textual analysis as its core approach. Through systematic close reading and critical analysis of the documentary films' system of signification and logic of spatial construction, this study attempts to provide an interpretation of documentary films. The adoption of textual research method is based on the epistemological position that documentary films are complete semiotic systems and carriers of meaning. Such an epistemological position demands that the researcher should go beyond simple description of film contents and explore the productivity of form itself, that is, how the audiovisual language through its specific organization produces certain spatial perceptions and emotional effects by itself.

In terms of operation, this study builds a two-dimensional analytical framework based on structuralist narratology and film semiotics. The analytical framework builds on the combinatorial relations of concrete technical aspects in audio and vision. The visual dimension analyzes the combinatorial patterns of stylistic elements such as shot scale, camera movement and lighting/color. The auditory dimension analyzes the rhetorical functions of elements such as voice-over narration, sound effects and music. The establishment of analytical framework enables the analysis of the abstract concept of “audiovisual construction” to rely on concrete and perceptible technical aspects, and thus provides a systematic way to analyze texts.

In terms of research texts, this study follows the principles of typicality and representativeness and selects mainstream rural documentaries produced since 2012 that have had a wide social influence as samples. The study takes authoritative series such as *Remembering Nostalgia*, *Beautiful Countryside*, *The Field of Hope: By the Lalin River*, *Beautiful Village* and *Searching for Craftsmanship* as main samples. These works are authoritative in terms of broadcasting platforms. They have distinct practices in exploring the audiovisual language, and together can reflect the overall landscape and aesthetic features of rural narratives in the new era.

The process of textual analysis follows the principle of progressive analysis, from surface meaning to underlying meaning, and from form to significance. First, it involves detailed viewing and description, and then records representative audiovisual segments and their technical characteristics accurately. Second, the observations are organized in the analytical framework to analyze the combinatorial relations and internal grammar among different audiovisual elements. Finally, based on the formal analysis, it conducts interpretation analysis to explore the rhetorical intentions and cultural logic embedded in specific audiovisual structures, and analyzes how they achieve emotional mobilization and meaning construction through providing sensory guidance. This study explores a layered analytical approach to establish an analytical chain from technical form to cultural meaning, and provides a systematic textual basis to explore the rural visual narratives of the new era.

2. Visual Strategies: How to Make “Messages” Visually “Moving”

As we have stated above, we are dealing with documentary texts, not the real world. Therefore, the spatial narrative that is being examined here is not the spatial narrative that is built in the world, but the narrative that is built in the filmed image. How do documentaries build space in the visual frame? What techniques are used to build this space?

It should be noted that there is often confusion in regard to the study of spatial narrative in visual media and in literature. Narrative in literature is based on “words,” while film and documentary do not have the idea of the “word.” As Pier Paolo Pasolini (1984) stated, “There is no dictionary of images. There are no images that are ready to be used. If we were to imagine such a dictionary, it would have to be an infinitely large one... The film director has no dictionary; he has nothing but infinite possibilities.” What then is the language of film? Knowing the answer to this question is the key to knowing the language of documentary used in the visual construction of space.

The well-known French film critic and theorist André Bazin, in his book *The Evolution of the Language of Cinema* (1980), studied and conceptualized “film language.” According to Bazin, “visual composition,” “montage techniques,” “cinematography and editing,” “image and sound,” “deep focus,” and “shot breakdown” are the basic elements of the language of cinema. Film and documentary use this basic language primarily to create narrative and, ultimately, certain artistic effects.

Indeed, the art of the moving image is inherently dependent on both “seeing” and “hearing.” The language of film and documentary encompasses a variety of visual and auditory methods. The spatial narrative in documentary is thus built upon the foundation of the visual image and sound. This analysis will proceed in two parts: first, how the documentary’s visual image constructs screen space, and second, how documentary uses sound to shape that screen space.

Our earlier discussion noted that film language includes commonly used, conventional techniques like “visual composition,” “montage,” “cinematography and editing,” “image and sound,” “deep focus,” and “shot breakdown.” However, film language is not a natural language. Jean Mitry (1988) contended that film expresses meaning directly through itself; therefore, speaking of its language, utterance, or concepts is merely metaphorical. As a form of direct signification, film is cognitively metaphorical because cognition involves the distinction between expressive thought and the object of thought. Conversely, scholar Warren Buckland suggested that film can be considered a language, or at least a grammar, noting that a shot is somewhat analogous to a word, and linking shots through editing resembles forming syntax.

Thus, the film language system constitutes a rich system of signs and coded meanings, whose organization and application are highly diverse. Building upon conventional film language, cinema develops distinct languages that correspond to specific film themes, styles, and discourses. For instance, film language manifests in concrete and unique ways according to a film’s theme, style, and discursive framework.

Michel Foucault's (1986) definition of discourse is very useful here. He defines discourse as meaningful statements made in certain historical periods, under certain historical conditions, according to certain rules and practices. Foucault defined the term "discourse" as a collection of statements that offer a kind of language, a way of talking about or representing something (Foucault, 1986: 94).

It is necessary to introduce the relationship of "film and discourse" because the Chinese countryside films in this new era to some extent possess the features of what Foucault once called "discourse". Entering a new era, the Chinese countryside has undergone tremendous changes and a new look. These new-era villages are new and beautiful, post-poverty alleviation villages, and they have characteristics of "goodness, beauty and truth". In order to produce the discourse of "goodness, beauty and truth", the documentaries need certain visual and auditory forms to create the linguistic syntax suitable for them.

As a visual art form, documentary forms spatial imagery through its visual language. In contemporary Chinese rural documentaries, the visual artistic language is complicated and diverse, but it tends to be a discourse of "beauty" overall. This beauty includes the beauty of space in terms of grandeur and aesthetics, the beauty of social space in terms of order and harmony, and the beauty of spiritual and cultural space in terms of emotion and intellect. In order to produce images of these three types of beauty, documentaries create their features of "beautiful space" and "virtuous/beautiful space" through their visual languages, such as shot scale, light and color, and camera movement.

2.1 *The Spatial Dialectics of Shot Scale: From "Grand Narrative" to "Micro-Emotion"*

Shot scale is one of the basic visual means for spatial narration. As Marcel Martin (1980) puts it, the shot is the most elementary unit of the language of films — the material of cinema. It is both the objective reproduction mechanism that mechanically recreates the image and an important manifestation of the director's intention. These cinematic elements include many aspects, and shot scale is of particular significance.

Be it in literature or in images, any subject can be depicted from its scale and scope. In cinematic arts, shot scale refers to the area and spatial range that a scene or character takes up in a frame. In general, documentaries use extreme long shots, long shots, medium shots, close-ups, and extreme close-ups as cinematic frames, presenting scenes or characters in different spatial ranges within the frame. And the choice and use of shot scale is closely related to the documentary's discursive context.

In contemporary Chinese rural documentaries, shot scale is a routine language for visual space construction, and full shots present a more overall picture of the spatial landscape of new-era countryside. Moreover, in some documentaries, certain shot scales become defining stylistic elements in the visual language of documentaries, shaping its visual style and discourse through constraining the use of shot scale.

The documentary *Remembering Nostalgia* uses extreme long shots and close-ups as defining stylistic elements. These two shot scales are polar opposites in cinematic language. The extreme long shot means the largest range of depicted space — the environment setting is the absolutely overlord element of the image. People almost disappear in the frame, becoming just color patches or lines before the grand view. While the close-up means the compression of the depicted visual space, magnifying and highlighting certain specific parts of scene or characters, almost excluding the environment setting of the scene.

Remembering Nostalgia uses a lot of extreme long shots to present the picturesque visual spaces of China's new-era countryside and arouses the audience's deep nostalgia. The director uses many extreme long shots and innovative aerial filming. Aerial filming means filming the scene from the sky or a high position, creating a panoramic view from a distance. At the same time, it also creates a view of omniscient bird's eye view, showing the overall geographical layout of villages and homelands.

When shooting Chakeng Village, the documentary uses aerial shots to show villagers drying tangerine peels in the courtyards. The whole frame becomes a painting. The aerial shot gives the audience a very poetic visual experience (Figure 1).



Figure 1. Drying Tangerine Peels in Chakeng Village from *Remembering Nostalgia*

The long shot not only presents a real scene but also can evoke people in Chakeng Village to remember the places in their childhood when they have left the village. In *Remembering Nostalgia*, the director uses long shots and empty shots (no people, only the scene) to create a mood. For example, a shot of a “sea of clouds” over Maihuayu Village, the camera shows the fog gradually dissipating and revealing the shape of the village clearly to the audience. This way, the visual space appears to be bigger. By showing these wide views, the film tries to display ecological environment beauty of China new-era countryside.

The long shot can greatly expand the rural space on screen, and long shots make the feeling of nostalgia more profound.

Remembering Nostalgia also uses many close-up shots. These shots try to reveal the emotional world in rural stories. Close-ups are completely different from long shots. By moving the camera close to the subject, close-ups focus on showing facial expression, hand gesture, or the texture of an object. This way, close-ups are more effective at evoking the audience’s emotions.

In the episode about Taoyuan Village, the director uses close-ups to show a girl’s feelings. The story is about entrepreneur Wu Xiaobo. When Wu talks about the difficult time when he started his business in his peach orchard, a series of close shots on his tear-filled eyes and his hand wiping the tears away are shown. These close-ups not only show that Wu feels guilty for returning to his hometown to start his business for his mother, but also make the audience deeply sympathize with him. The emotional distance between the close-up and the viewer becomes very short. The personal feelings of Wu are feelings that the viewer also has.

These close-ups are important to tell this story. Wu Xiaobo’s quick thinking and knowledge help him become the first person in the village to sell peaches online. At this moment of success, the director uses a close-up on Wu’s eyes, which are now filled with happy tears, as Wu says, “I am still my mother’s pride.” These close-ups used at the low and high points connect the different parts of Wu’s life and show Wu as a brave and capable new countryside person.

Documentaries about ICH are especially good at using close-up shots. With the development of China, people’s interest in culture traditions gradually increased. This has led to many documentaries about rural ICH, such as *The Great Hidden Craftsmen*, *Searching for Craftsmanship*, *Guardians of Chinese Crafts*, and *Heritage at Your Fingertips*. These films show heritage keepers and their crafts. Close-up shots are very important to show the dedication of artisans and the texture of crafts.

For example, *Searching for Craftsmanship* presents crafts such as dough figurines, pottery, and papermaking. When introducing these crafts, close-up shots are indispensable. Close-up shots allow viewers to see the details of the craft being made and also capture the clear trajectory of the artisan’s movements, their expression of concentration, and their intense gaze (Figure 2 and Figure 3).

In total, the scale of shot helps create the visual space of the film. The use of close-up shots in the new-era Chinese rural documentary is very appropriate, and it also closely corresponds to the cultural theme of the new era.

Figure 2. *Searching for Craftsmanship*Figure 3. *Searching for Craftsmanship*

2.2 The Spatiotemporal Sculpting of Light: From “Physical Illumination” to “Spiritual Revelation”

Light is an important visual means of spatial narration. It is an essential component of the composition, creating different effects in the space within the frame. First, light is closely related to screen space. Light appearing in the visual frame gives emphasis and definition to the spatial environment. In Western classical painting as early as in the 16th century, light has been used to create spatial perspective and depth on the two-dimensional canvas. After the invention of film, light still bears the same function. For example, zonal lighting can break the uniform illumination, increasing the spatial depth of the image (Wang Juren, 2022). In documentary filmmaking, using light is also indispensable. Creators shape three-dimensional and volumetric space by using lighting to transcend the two-dimensional plane.

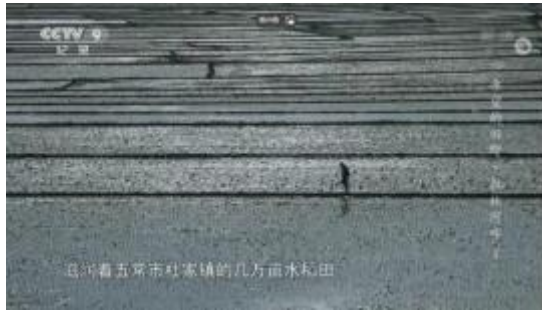
Second, the relationship between light and color should be closely examined. Light in cinematic arts can be divided into natural and artificial lights. The sun is a source of natural light; the color temperature of natural light varies with the time of day. Artificial light is man-made; it can emit lights of different colors. In film and documentary, the use of natural light and artificial light creates diverse and colorful images.

Finally, the relationship between light and shadow cannot be ignored. Light falls on the plane and forms bright areas; the shadow part is removed by light. This lighting ratio, the contrast between light and dark, models the three-dimensionality of the image and enhances the spatial effect.

The main topic of the documentaries in the new era is the new era itself, that is, protected and developed ecological environment, harmonious co-existence of man and nature, comfortable and suitable living environment, and more comprehensive cultural prosperity. In this regard, light, as an important element of cinematic art, plays a vital role. In the process of shooting documentaries, creators use light for modeling, which is not only essential for creating the overall atmosphere and emotional feeling, but also for modeling the visual space of the image.

Documentaries use light to reflect and represent the tranquility and harmony of the new Chinese countryside. In the process of filming, cinematographers use natural light and shadow effects to create the atmosphere of the visual space. For example, in the documentary *The Field of Hope: By the Lalin River*, it records the rapid development of Dujia Town in Wuchang City, Heilongjiang Province, known as China’s “Land of Rice” (Figures 4 and 5). Through depicting the prosperous life of the people in Dujia Town, who live on the Wuchang rice nourished by the Lalin River.

During the *Grain in Ear* period, it is seen that the Lalin River sparkles in the sunlight, irrigating the vibrant green rice seedlings. In the harvest season, the golden rice fields undulate like waves under the sun. In winter, Dujia Town is covered with snow; the sunlight reflects on the snowy ground, making the vast land look majestic and splendid.

Figure 4. *The Field of Hope: By the Lalin River*Figure 5. *The Field of Hope: By the Lalin River*

The series *The Best Time of the Year* broadcast by Hunan Satellite Television records how the villages in Hunan Province had been lifted out of poverty and became *beautiful countryside*. The eight episodes of the series authentically reflect the beautiful landscapes and rich cultural heritage of Hunan's countryside. The simple and naturally beautiful lives of villagers are presented in the series.

There are many scenes reflecting the beautiful sceneries of the countryside in Hunan in the documentary. For example, the morning sun shoots through the mist and illuminates the fields; an old farmer works on the land. Light and shadow make shadows on the face of the old farmer telling a story of his hard work during his lifetime. All these reflect the beauty of the countryside in Hunan and the happy lives of villagers.

In addition, documentaries use light to reflect the historical and cultural spiritual space of the new-era Chinese countryside. Compared with urban documentaries, not only can rural documentaries use light to reflect three-dimensional space and visual beauty of rural life, but also use light to reflect psychological emotional space and reflect the spiritual space of rural history and culture. Documentary filmmakers shoot light into the lens and audience sees the light and objects illuminated by light. Xiang Xianghu said that "human consciousness can impose an a priori form on the sensory field, grasping the visual structure as a whole" (Xiang Xianghu, 2018). Visual beauty can stimulate psychological response, arousing space of memory, emotion and spirit. It is the same in shooting rural documentaries with the theme of nostalgia.

Taking *Remembering Nostalgia* and *Chinese Villages* as examples, using light to shoot ancient historical buildings in the countryside reflects beautiful sceneries. Not only can it construct the architectural and structural space of the village in the frame, but also it can arouse the psychological response of audience. Arouse memory, emotion and a spiritual space of specific histories. I believe that in shooting nostalgia-themed rural documentaries, filmmakers successfully use light to connect the two-dimensional visual space of documentary and the psychological space of audience. That is, viewing is no longer a two-dimensional visual experience, but a connection of internal space of documentary and mental space of audience. It breaks the "fourth wall".

In *Remembering Nostalgia* and *Chinese Villages*, the director intentionally plays with the light and shadow effects to display the beauty of ancient historical building and landscape in the countryside (Figures 6, 7, 8, 9, 10, 11). These old villages and buildings are "static" and "silent". But to give them "vitality" and "energy", to make them "come to life", to record the documentaries the light that falls on these villages and buildings for a period of time will certainly be "flowing space". This "flowing space" makes these villages and buildings appear three-dimensional and alive. This "come to life" visual effect will certainly trigger an internal response in the audience. They will trigger audience to imagine walking in the space unconsciously. For those who once residents but have left behind will have a special affection and magnetism for these images. They will trigger memories of history and old years and attract them to enter in this emotional space of "nostalgia".

In a certain sense, the old Chinese countryside should release its internal driving force. Even these old villages and buildings that represent old history and culture, if the flow of light and shadow helps them to become a dynamic and living space in the image of the new era.

Figure 6. *Remembering Nostalgia*Figure 7. *Chinese Villages*Figure 8. *Chinese Villages*Figure 9. *Chinese Villages*Figure 10. *Chinese Villages*Figure 11. *Chinese Villages*

2.3 The Spatial Weaving of Moving Cameras: From “Static Observation” to “Dynamic Immersion”

Composite shots with multiple angles and directions are an important way to express space in rural documentaries. As Jiao Xiongping pointed out, Louis Giannetti uses different English terms such as “Movies,” “motion pictures,” and “moving pictures” in his book *Understanding Movies* to define the form of camera movement. Both of them show that moving pictures have rich meanings of motion and essential meaning. René Clair once said that movement was “the very substance of film esthetics.” Thus, motion is its nature and makes a distinction with other arts.

Unlike static shots, moving shots can change the optical axis of the camera and the position of the camera or even the lens angle to realize the movement such as dolly-in, dolly-out, pan, track, follow, crane-up and crane-down. In fact, these kinds of movements are only the simulation of human’s own way of observing. That is to say, they just simulate people’s behaviors of approaching, distancing, following, panning, and changing viewpoint and direction. Therefore, different kinds of movement forms would realize different visual effects. That is, one moving shot can present one certain space or connect several spaces together; moving shots often do not realize only one kind of movement in one take, that is, they would appear as composite shots with two or more kinds of movement forms in one take. These kinds of composite shots would realize rich viewpoint changes, that is, change of angle, direction, and height, and allow people to observe space in full view. Simply speaking, as audience watch different kinds of elements in one scene of documentary, they would realize space from the motion of camera and changing of several viewpoints. In addition, different kinds of movement forms would realize different kinds of visual spatial effects and atmosphere. For example, dolly-in would realize tension, excitement, or concentration; dolly-out would realize breadth, freedom, or relaxation; and pan would realize a rotating, dynamic, and lively visual effect.

First, overall, moving shots with multiple scales, angles, and directions would realize stronger expressive power and deeper level of spatial narration in new-era rural documentaries. On the one hand, they would allow people to better realize the spatial layout and environment of countryside. For example, flying over fields and orchards or following streams would realize geographical features and natural environment of countryside. On the other hand, they would allow people to record and realize activities of villagers. Following villagers who come back from working in fields or children playing at the village entrance would allow people to better realize life of countryside. In addition, moving shots would also allow people to realize different speeds of movement. Different speeds would allow people to realize countryside at different rhythms. For example, slow tracking would realize tranquility and peace; fast moving would realize vitality and change. They would also allow people to realize and speed up or slow down documentary. For example, when documentaries focus on seasonal changes, they would allow people to realize process of growth, flowering, and fruiting of plants. It would allow the documentary to realize more lively and vivid narrative pace. In addition, they would allow people to realize documentary in different visual effects and make documentaries more attractive. For example, using drones to follow and shoot fields and landscape from the point of view of animal would allow audience to realize different kinds of visual effects.

Second, moving shots can connect disparate, individual spaces into a cohesive whole through specific movement patterns. Within the visual frame, this helps viewers form an overall impression of the filmed space. Unlike perceiving real-world space, the space presented within a shot is always limited. Moving shots, however, can traverse from one space to another, or to multiple spaces, presenting space panoramically within the cinematic world.

Take the three-episode series *Beautiful Countryside* from Zhejiang Satellite TV as an example. It tells the story of Gejia Village in Ninghai, Zhejiang, where resident artists used art for rural construction activities. It vividly shows how the artists, “using the village as a platform and art as the medium,” transformed the countryside, shaped and depicted its beauty, and contributed to rural revitalization. Divided into chapters titled *Artists Came to the Village*, *Villagers Became Artists*, and *A Beautiful Countryside Today*, the creators made extensive use of moving shots. These shots follow the villagers of Gejia Village as they go about their daily productive lives and activities within their familiar rural spaces, while also showing them renovating their own village, inspired by the resident artists. These moving shots give the audience a comprehensive impression of Gejia Village’s overall spatial layout, introducing spaces that were, or had the potential to be, transformed. The use of moving shots in the documentary resonates with the words of resident artist Cong Zhiqiang: “For rural revitalization to be truly achieved, villagers themselves must become active participants.”

Moving shots can follow characters through space, allowing the audience to also perceive the width, length, and depth of the environment alongside them. For example, following a person, creating a subjective point-of-view shot, can give the audience a sense of spatial immersion and experience. In the first episode of *Beautiful Countryside*, *Artists Came to the Village*, which primarily depicts artist Cong Zhiqiang’s initial arrival in the village, the cinematographer uses shots from multiple angles, directions, and movement types to show Cong exploring every corner of the village on foot (as shown in Figures 12, 13, 14, 15).



Figure 12. *Beautiful Countryside*



Figure 13. *Beautiful Countryside*

Figure 14. *Beautiful Countryside*Figure 15. *Beautiful Countryside*

3. Auditory Strategies: How Does Sound “Construct” an Ideal Countryside?

As any sound recordist knows, when a microphone and recorder are set up in a supposedly ‘quiet’ environment, a multitude of previously unnoticed sounds inevitably emerge, precisely because the microphone is non-selective (Bordwell, D., Thompson, K., & Smith, J., 2010). This indicates that an absolutely “silent” space does not exist in the real world. As a spatial art, film and documentary rely not merely on the medium’s capacity to record sound, but more fundamentally on sound itself being an integral component of space. A space without sound is incomplete and cannot be fully realized (Chen Gong, 2016).

In addition, extending analysis of space to the auditory dimension, the appearance of the “soundscape” substantially enriches our knowledge of the “sound-space” relationship. The soundscape was defined by R. Murray Schafer as follows: “an acoustic environment as perceived by an individual or society, focusing on how we perceive and conceive it”. The term “scape” is always linked to “seeing”, and therefore has imagery and visibility. How can sound be seen? In fact, when a certain sound appears in our ears, we can imagine an image of the sound, and then, according to the image, we can draw a concrete spatial image according to our experience (Chen Yan, 2015).

After the introduction of the concept of the soundscape, it is widely used in fields such as sociology, art studies and acoustics. Based on the introduction of the soundscape theory, the film soundscape concept has gradually been proposed. What is the film soundscape? In general, it is the sonic landscape in the film. Its connotation is reflected in the following aspects: first, the objectively recorded sonic documentation; second, the holistic consideration of sound in the process of sound design and musical composition; third, the spatially embodied and immersed experience and aesthetic perception in the film’s audiovisual aesthetics; fourth, the relationship between the sonic environment and people in the film, as well as the structural relationship between the film sound and the audience (Hou Dandan, 2023). The concept of the film soundscape fully recognizes the important position and role of sound in the expansion of visual space and immersion.

In the new-era Chinese rural documentary, sound also plays an equally important role, so the analysis of its auditory construction method is especially important. However, it should be noted that the narrative focus of the new-era documentary is fundamentally different from that of the previous documentary. Not only tell China’s stories well and speak for China, but also, through a creative model of “voice-over narration + music”, narrate the stories of ordinary people in the Chinese countryside, display the simple beauty, cultural connotation and national feeling of *Chinese villages*, but also actively introduce China to the world.

3.1 Voice-Over Narration: From “Informational Commentary” to “Spatial Navigator” and “Emotional Resonator”

The use of voice-over narration is quite common in rural documentaries. It is a written text, created by the director at the stage of scripting to describe the image, read from the camera and then recorded as a voice. Functioning as commentary, the voice-over explains and supplements the visual. As an important part of the soundscape, the voice-over assists the audience in forming an auditory space outside the visual. Not only does it enhance the audience’s understanding of the space on the screen, but also triggers their imagination of the space outside the screen through sound. As a result, the voice-over plays a vital role in diminishing the audience’s dependence on the visuals themselves, enhancing their perception of space both inside and outside the frame through a whole space outside the screen, and assisting them in forming a clear idea of the total environment.

First, the use of voice-over narration crafts an experience of the realistic space within the rural visuals for the audience. In documentaries, voice-over serves important narrative functions: it works in concert with the images to drive the plot and supplements the visual information. Through voice-over, viewers can achieve an “immersive” spatial experience. When watching rural documentaries, the combination of imagery and

voice-over allows for a more authentic perception of the presented space, creating a strong sense of “being there.”

In the fourth episode, “*Gazing at the Hometown*,” of the new-era documentary *Chinese Villages*, the creators used voice-over to depict the salt-drying practices of Tibetan people in China. The narration described: “Every spring, in the Grand Canyon of the Lancang River, groups of women can often be seen walking briskly. Setting out from home early in the morning, they make countless trips back and forth throughout the day. Their task is to carry brine from the salt wells, bucket by bucket, to the salt fields to be dried into salt...” This commentary aligns perfectly with the visuals. Its use of numerous action verbs makes the salt-drying process more tangible and vivid. The audience not only gains a deeper understanding of the on-screen action but is also drawn in by the evocative description, feeling as if they are personally participating in the salt-drying process and immersed in the authentic living space of the Tibetan people.

Just as in the above part of documentary about daily life of Chinese people in *beautiful villages*, the voice-over in *Beautiful Countryside* also describes the life in China in very expressive and poetic ways. For example, the voice-over in *Beautiful Countryside* describes the life in countryside as: “On the grass land under green trees, there are children running and playing with happy smiles, their happiness spreads to the whole village; next to the stove that is burning fiercely, there is an old person sitting quietly, smoking a pipe and enjoying this peaceful and happy moment”. Through the description in voice-over, the audience can imagine the scene as if they are really in the countryside of new era and enjoying the leisure and happiness. At the same time, the expressive voice-over also builds an ideal space of the countryside and triggers audience’s desire for a poetic life.

Second, as mentioned above, the voice-over not only can supplement the information of visual and also let the audience immerse in the on-screen space, but also can let the audience enjoy the description with feelings and trigger their empathy. Through the voice-over with rich feelings and reflecting the feeling of audience’s staying in the on-screen space, it triggers audience to immerse themselves in the emotional space beyond the on-screen space.

New-era rural documentary is different from other rural documentaries. It focuses on the feeling that Chinese people have for the countryside. The voice-over with feelings, reflecting the feelings that the creators have for the countryside, can more effectively trigger the audience to immerse in the emotional space beyond the on-screen space. In short, in addition to the experience of space in the rural visuals and drive the plot forward, the voice-over also on a higher level triggers audience to immerse in the emotional space beyond the visuals.

In *Beautiful Countryside* about “Beautiful Countryside Construction”, the voice-over has strong feelings. For example, in the *Beautiful Countryside* about *North of the Pass*, there is a voice-over as follows: “The once “Great Northern Wilderness” has now become the “Great Northern Granary”, but the black soil still bears the pursuit of happiness and the hope for future of generations of people living in this place”. This voice-over reflects the creators’ reflections on the huge change made by the development of countryside in Northeast of China from being a wasteland to a granary. This change brings both happiness and a sense of future. This kind of voice-over not only reflects the hard work of the people in Northeast China, but also prompts the audience to imagine the region’s developmental process and evokes emotional resonance.

In *A Bite of China*, the narration states: “On these open fields, spring blossoms to autumn fruits, and the cycle of life goes on. Every seed contains the power of life; every harvest is filled with gratitude and happiness. This gratitude and reverence for life that comes from our hearts towards nature”. This kind of voice-over with feelings not only strengthens the audience’s immersion in the on-screen space, but also triggers them to enter the emotional space beyond it.

3.2 Sound Effects: From “Environmental Background” to “Spatial Authenticator” and “Ideological Loudspeaker”

Sound effects is one of the elements of spatial narrative in rural documentaries. Sound effects is an important part of film sound and film soundscape. In general, sound effects are the noises made by objects in the frame by movement, collision or impact. Sound effects can be further divided into two categories: natural sounds and mechanical sounds. Natural sounds are all acoustic phenomena that can be heard in nature but are not produced by living things. Natural sounds include natural phenomena such as wind, water flow, rain, thunder and dripping water, as well as animals such as cats, dogs, crows, bees, frogs, insects, cows, crickets, owls and cicadas. Mechanical sounds are noises made by human movement or machines, including human speech, sounds made by machines such as radios, and special effects such as door opening and closing, vehicle sounds, crackling flames, footsteps on snow, wind chimes, temple bells and creaking water wheels. Natural and mechanical sounds make up the basic elements of film soundscapes.

Hou Dandan (2023) finds that films mention the clear differences in soundscapes between villages, towns and cities. She also writes that with urbanization, many villages have been swallowed into the outskirts of cities, and

the original soundscapes of these places are becoming harder to distinguish from the city soundscape. Additionally, similar models of urban development have made the soundscapes of these villages less regional and authentic.

Cities and villages are two distinctly different kinds of space in the real world, and this difference is also reflected in the sounds. In cities, the sounds of machines can be heard — people's voices, machinery screaming, and houses under constant construction. In villages, the main sounds are natural — the rustling of rice and leaves, the swaying of bushes, and insects and birds chirping. So, in new-era Chinese rural documentaries, these real rural sounds are used to really set the authentic countryside atmosphere.

First, sound effects in rural documentaries enhance the three-dimensionality and spatial perception of the visuals. As argued, sound significantly increases the fidelity and credibility of the image, not just materially but also aesthetically. In rural documentaries, the high recognizability of sound effects markedly boosts the audience's spatial awareness. Creators integrate these sounds with the visuals, making the depicted space feel more tangible and authentic.

The documentary *Beautiful Countryside* uses sound effects to show the environment in areas with sound effects of rural fields in large quantities and makes people feel immersed in the environment when they watch the documentary. In portraying characters, gentle music and tender lyrics may reflect gentle and weak personalities, and nervous music and hasty and hurried sound effects reflect anxious and nervous mood. These sound effects enhance the realism of space in addition to the image.

Similarly, *Searching for Craftsmanship* pays meticulous attention to sound when showcasing traditional crafts. It carefully records the specific sounds of the creation process—the knife cutting wood, the hammer striking metal, the crackling fire during pottery firing, the clang of worked iron and copper, and the grinding of blades. These sounds allow the audience to accurately perceive the spatial environment of the workshop and feel immersed in the hands-on process of craftsmanship.

Second, as a rhetorical device, sound effects help to imbue the new-era countryside with mainstream ideology.

Jacques Attali (1985) writes of the opposition between noise and music: Noise gives rise to chaos; music gives rise to power and subversion. Noise, if invested with particular tools and turned into sound, yields power, purpose and dreams — the very stuff of music. Attali places music in the midst of a rationalizing aesthetics, but also uses it as a haven of the irrational. He situates music as power and as play.

This perspective aligns with analyses of China's New Mainstream Cinema. Scholars note that these films utilize sound to express relationships with mainstream ideology and values. For instance, in *The Silent War*, the incessant ticking of telegraph signals evokes the mystery and peril of undercover work. In *The Founding of an Army*, the unified, rhythmic marching of troops sounds tense yet orderly. In *The Hundred Regiments Offensive*, the rumbling artillery fire conveys the intense brutality of war.

In the new era, in addition to displaying rural life, Chinese rural documentaries also bear the responsibility of promoting national mainstream ideology and social values. And it is precisely through the effective use of sound effects that these documentaries express their specific spatial discourse, meeting the new thematic requirements.

In *Beautiful Countryside*, one important theme is to display the local customs and natural scenery of China's villages. To create an atmosphere of picturesque scenery and a harmonious social environment, the sound recordist deliberately records various typical sounds in nature and in the countryside—the singing of insects and birds at dawn, flowing water, rain, and wind. The combination of these natural sounds audibly constructs the picturesque natural space of new-era Chinese countryside.

Similarly, in *Remembering Nostalgia*, which intends to display the hard-working, simple, and beautiful Chinese peasants and their inner world, the creator also records sounds of people talking and laughing cheerfully, the sounds of daily life ambience, and the sounds of labor.

In summary, in documentaries reflecting the development of Chinese countryside in the new era, in addition to creating atmosphere and expressing emotion, sound effects also bear the function of constructing ideological space for the new-era countryside. This function is of great significance to the thematic imperatives of the new era and the spatial discourse of the present, and it may also be reconstituted and passed down as collective memory in the future.

3.3 Integrated Soundscape: From “Elemental Overlay” to “Acoustic Archive of Collective Memory”

Voice-over narration and sound effects do not exist in isolation; they function synergistically. The narration assigns meaning to the sounds (“This is the sound of harvest joy”), while the sound effects provide texture and verification for the narration. Together, they weave a complete and persuasive soundscape.

The ultimate aim of all auditory techniques is to collectively construct the documentary as an “Acoustic

Archive” of the new-era Chinese countryside. This archive selectively collects and showcases those sounds that align with contemporary discourses such as “Beautiful Countryside,” “Rural Revitalization,” and “Cultural Confidence.” These ritually presented sounds are not merely records of the present moment. They are designed to sediment into the collective auditory memory of the future, perpetually shaping how people perceive and imagine the Chinese countryside of this era.

4. Conclusion

This study systematically examines the visual and auditory construction mechanisms of spatial narratives in contemporary Chinese documentaries focusing on rural themes. The findings reveal that through visual techniques—including shot composition, lighting, and camera movement—these documentaries not only depict rural physical environments but also construct a multidimensional space integrating geographical landscapes, psychological emotions, and cultural memories. The dialectical application of long shots and close-ups creates a dynamic interplay between macroscopic narratives and microscopic insights. Lighting transcends its basic function of physical illumination to become a spiritual element that sculpts temporal dimensions and reveals historical connotations. Composite moving shots simulate embodied experiences, facilitating a transition from static observation to dynamic immersion in spatial perception.

On the auditory level, voice-over narration and sound effects collectively form a comprehensive system for “spatial construction through soundscape.” Voice-over serves dual functions as both spatial navigation and emotional resonance, guiding viewers beyond visual frames into expanded emotional and imaginative spaces. Sound effects function not merely as environmental background but as authenticating codes for spatial authenticity and subtle vehicles for ideology, collaboratively shaping rural imagery that aligns with mainstream value expectations.

Visual and auditory elements are not employed in isolation; rather, they interpenetrate and mutually reinforce each other, constructing a highly integrated system of “spatial rhetoric.” Ultimately, through this complete audiovisual syntax, contemporary rural documentaries successfully transform rural space from a mere geographical entity into a symbolic field that carries national discourse, zeitgeist, and cultural confidence. Their spatial narratives not only achieve aesthetic innovation but also profoundly participate in shaping collective memory about the modern transformation of “rural China” and in conveying ideological messages, thus establishing themselves as significant cultural practices for narrating China’s stories and transmitting China’s voice.

References

- Attali, J. (1985). *Noise: The political economy of music*. Manchester University Press.
- Bazin, A. (1980). The evolution of film language. *Film Art Review*, (2). (Translated by Cui Junyan & Xu Zhao).
- Bordwell, D. Thompson, K., & Smith, J., (2010). *Film art: An introduction* (Vol. 7). McGraw-Hill.
- Chen, G. (2016). The design principles and methods of ‘spatial sense’ in film sound. *Film Art*, (1).
- Chen, Y. (2015). On the composition of film space narrative. [Master’s thesis, Nanjing University of the Arts], p. 9.
- Cong, Y. (2019). Research on the rhetoric of film language in the New Era’s Main Melody movies (2012-2018). [Master’s thesis, Hunan Normal University], p. 32.
- Foucault, M. (1986). Of other spaces (J. Miskowiec, Trans.). *Diacritics*, 16(1), 22-27.
- Giannetti, L. (2013). *Understanding film* (11th ed.). Beijing World Publishing Corporation, p. 54.
- Gu, Y. B. (2002). Music and sound in film and television art. *Contemporary Communication*, (4), 92-93.
- Hou, D. D. (2023). From film soundscape to soundscape film: Exploring film sound based on soundscape theory. *Film Literature*.
- Martin, M. (1980). Film language [Doctoral dissertation]. (He Zhengan, Trans.).
- Mitry, J. (1984). *Jean Mitry and the aesthetics of the cinema*. Umi Research Press.
- Pasolini, P. P. (1984). Citation details required — original source information incomplete.
- Schafer, R. M. (1993). The soundscape: Our sonic environment and the tuning of the world. Simon and Schuster.
- Wang, J. R. (2022). The expansion of film visual space through digital lighting. *Contemporary Film*, (4).
- Wang, X. Y., & You, Y. W. (2024). Thematic expression and audiovisual presentation in documentaries on rural revitalization. *Audio-Visual*, (10), 17-20.
- Wang, Y. W., & Yang, S. (2024). Cultural tourism variety shows: Audiovisual representation and value reconstruction of modern rural space. *China Television*, (10), 51-55.

- Xing, X. H. (2018). From image to film: Suturing art theory based on visual perception psychology. *Contemporary Film*, (7).
- Zhang, H. Y. (2024). Chinese experience of digital audiovisual communication assisting rural revitalization. *Media Observation*, (5), 1-12.
- Zhao, L. (2024). Audiovisual representation and value reconstruction of modern rural space. *News Gathering and Editing*, (5), 33-34.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).