Distinctive Sound Design—A Comparative Study on Audiovisual Analyses of Different Types of "Immersive" Short Videos

Jiayi Wang

Communication University of China, Beijing, China

Abstract: *This paper constitutes a comparative investigation into various genres of "immersive" short videos, with a primary emphasis on the comprehensive analysis of the audiovisual language employed therein. The objective is to discern the distinctive features embedded in the sound and image components, subsequently delving into the singularly designed auditory elements characterizing this particular genre of short videos. The introductory segment of this paper provides a delineation and definition of short videos and specifically categorizes them as "immersive." Proceeding to the core of the comparative inquiry, the paper systematically synthesizes and contrasts the auditory and visual attributes intrinsic to diverse types of "immersive" short videos, drawing upon the typologies identified by prominent short video platforms. Furthermore, it undertakes an examination of the disparities between "immersive" short videos and their conventional counterparts, scrutinizing the production aspects in light of the outcomes derived from the comparative analysis. Building upon the findings of the comparative study, a detailed analysis of the production-level distinctions between "immersive" short videos and conventional counterparts is presented. The concluding section delves into an exploration of the substantial value that sound contributes to "immersive" short videos, concurrently addressing the potential dual-sided social impact of immersive media in the future. The overarching aim of this paper is to furnish insightful creative recommendations for new media producers, facilitating a more effective integration of immersive media into the fabric of future society.*

Keywords: Short Video; Sound; Immersive; Media.

1. INTRODUCTION

With the development of technology and the continuous evolution of media, the conventional modes of sharing images and text have become inadequate in meeting societal demands. Approximately in 2015, the introduction of short video platforms, such as TikTok and Kuaishou, marked the commencement of an era characterized by the consumption and generation of short video content, resulting in a rapid surge in short video users within China. By June 2020, the number of online video (including short video) users in China had reached 888 million [1]. Subsequently, the global popularity of the TikTok application in the following year solidified the maturation of the short video industry. Presently, the academic realm lacks a standardized definition for the concept of short video. According to the 2016 China Short Video Industry Research Report defines the concept of short video as "a video with a duration of tens of seconds to three minutes, with a maximum length of no more than five minutes, primarily relies on mobile intelligent terminals for rapid capture and aesthetically enhanced editing, presenting a novel video format that allows real-time sharing and seamless integration on social media platforms. Integrating text, voice, and video elements, it provides a more intuitive and three-dimensional means to fulfill users' expressive and communicative needs, addressing the demands for presentation and sharing among individuals. " [2]. Variances in the definition of short video exist among scholars, depending on the specific nuances and focal points of their research. This paper posits the definition of short video as follows: a video expression utilizing the Internet as a communication conduit, with a duration of less than 10 minutes, predominantly disseminated through mobile social media platforms.

2. "IMMERSIVE" SHORT VIDEO AS A SUB-GENRE OF SHORT VIDEO

The emergence of "Immersive" short video represents a recent evolution within the broader spectrum of short video. This sub-genre has swiftly gained popularity across major short video platforms, establishing itself as an autonomous topic classification with sustained momentum. In 2021, the series of "Immersive" short videos created by the TikTok account "Yangmaoyue" garnered widespread attention on the internet. The ensuing phenomenon, notably exemplified by the "Immersive Versailles" video series, sparked a trend wherein the notion of "everything can be immersive" captivated online audiences. Various inventive approaches to producing "immersive" short videos gained traction, rendering it an indispensable keyword in contemporary discourse. Platforms such as TikTok and Xiaohongshu witnessed a surge in engagement with diverse "immersive" themes, including

"immersive experience going home, " "immersive cake making, " "immersive makeup, " "immersive eating, " and "immersive unboxing, " boasting remarkable play counts and user comments. Despite the burgeoning interest in "immersive" short videos, scholarly exploration in this domain remains limited. Existing studies predominantly adopt varied perspectives such as content classification, as evidenced by the exploration of "beauty and makeup" [3], platform-centric analyses focusing on TikTok [4], and account-specific investigations exemplified by "Tao Lulu" [5]. This paper constitutes a comparative study into "immersive" short videos, scrutinizing their audiovisual attributes within the context of short video production. Through a systematic juxtaposition of different "immersive" short video types, we aim to analyze the distinctive audiovisual characteristics of this sub-genre and subsequently explore the underpinnings of its widespread popularity and social significance.

3.   THE COMPARATIVE ANALYSIS OF "IMMERSIVE" SHORT VIDEOS

This study focuses on meticulously selecting "immersive" short videos with notable engagement metrics, such as high clicks, plays, likes, and comments, across prominent media platforms both within China and internationally. The analytical approach involves deconstructing and scrutinizing these selected short videos to draw comparisons and distill the distinctive features of sound and image inherent in diverse "immersive" short video types. Subsequently, the obtained comparative results are synthesized and analyzed, shedding light on the unique design characteristic of "immersive" short videos. In the concluding phase, an in-depth exploration delves into the multiple values of "immersive" short videos, exploring their varied implications across aesthetics, business, and the humanities.

3.1 Uniqueness - Audiovisual Characteristics of Different Types of "Immersive" Short Videos

Contemporary mainstream video platforms employ diverse classification schemes for short videos, predominantly centered on thematic and content considerations. TikTok, for instance, categorizes short videos into 17 topics, Xiaohongshu into 10, and YouTube into 8. This paper directs its focus towards "immersive" short videos as a distinct subcategory within the broader classification of original short video content. Through a judicious combination of the existing short video content classification and a nuanced refinement of the thematic scope encompassed by "immersive" short videos, six distinctive categories emerge: operation-type, indoor programs-type, Vlog-type, study-type, sports-type, and funny-type. By comparing the audiovisual characteristics of the cases, the results of the study are reflected in the following five key dimensions.

3.1.1 Audiovisual Characteristics of Operation-type "Immersive" Short Videos

This category of short videos exhibits a wealth of sound and image. Broadly, the video content comprises intuitively realistic images, often characterized by a relatively raw picture quality. Certain videos within this classification maintain a cinematic aesthetic with the inclusion of subtitles. The auditory elements primarily consist of a blend of sound effects, music, ambient sounds, and human voices. However, the integration of these four elements does not follow a fixed pattern. The emphasis lies in aligning with the overarching theme of "operation", ensuring congruence with the core content expression of the video.

More specifically, the visual components frequently employ techniques to accentuate the central composition, enlarging specific sections and emphasizing details. Some images may embrace a gritty aesthetic, exhibiting a high degree of the grainy texture of video, while others lean towards an incomplete art style—occasionally extending beyond the screen boundaries, presenting a unique aesthetic perspective. The auditory features are diverse, predominantly falling into three categories. First, the sound includes music and sound effects, with an emphasis on the texture of the sound effects. The loudness of key sound effects is amplified during the production process, with other sound elements intentionally omitted. Music, often playing continuously with slight adjustments at key point, serves to highlight or obscure specific sound effects. Second, the sound includes sound effects and ambient sounds, especially in videos depicting immersive activities. These sound effects typically remain minimally modified, contributing to a sense of rawness and realism to enhance immersion. Third, the sound includes off-screen voice and sound effects, with a considerable portion of the off-screen voice being generated by AI or electronic vocals. Sound effects may either be individually and specifically recorded or directly selected from sound effect databases.

3.1.2 Audiovisual Characteristics of Indoor Programs-type and Funny-type "Immersive" Short Videos

This subcategory of short videos mirrors traditional TV program production in terms of both sound and image. Generally, video compositions are influenced by television program aesthetics, employing fixing seat shots and

interspersing individual close-ups or key shots strategically. In terms of sound, both voice and music are somewhat weakened, with music serving mainly to enhance the atmosphere and avoid dullness experience.

More specifically, the central visual elements of "immersive" short videos in this category are compact, relying on actions, behaviors, and objects to convey the narrative. In character-centric short videos, the number of characters on screen is reduced, often featuring only one person as the primary character. For object-focused short videos, physical objects within the frame assume a more prominent position, with filmmakers deliberately highlighting specific object details. Regarding sound, the significance of speech is diminished, and rather than conveying narrative content, human voices are compressed into concise expressions—only a nouns or verbs, and at times, even breathy voice.

3.1.3 Audiovisual Characteristics of Vlog-type "Immersive" Short Videos

The Vlog-type "immersive" short video, a trending topic in recent years, shares similarities with TV reality shows, thereby reflecting the influence of this television genre on its audiovisual style. The combination of sound and image in Vlog-type "immersive" short videos can be delineated in two key aspects. In terms of content and image, these videos tend to be shorter in length, adopt a more casual shooting style, and prioritize the recording of real-life experiences in a documentary fashion. In terms of sound, the human voice in Vlog-type "immersive" short videos is primarily recorded in real-time, either through direct recording or dubbing, with occasional use of AI-generated voices. The prominence of music is reduced, and some videos omit background music altogether. The amplification of ambient sound becomes a significant element, due to "immersive" short videos are keen to show real environments. For instance, in bustling restaurants, unembellished, noisy voices and continuous action effects capture the realism of the setting. Similarly, in quieter venues like homes or libraries, microphone background noise accentuates the quietness of the environment, contributing to an immersive viewer experience. Ambient sound and sound effects thus serve as crucial sources of realism and immersion.

3.1.4 Audiovisual Characteristics of Study-type and Sports-type "Immersive" Short Videos

Broadly, both study-type and sports-type "immersive" short videos share a common emphasis on showcasing the processes and states of learning and doing sports through video image, with speech not constituting the primary content. Specifically, the majority of images in these videos employ a fixed shot without extensive editing or collage, and there is minimal use of variable speed processing. Unlike variable speed processing, "immersive" short videos in these categories are more prone to leveraging editing techniques, maintaining the original speed of key segments while employing fast-paced editing techniques to stitch together and pile up them. At the auditory level, these short videos prioritize the preservation and even amplification of every detail of sound effects, especially those associated with key segments. For example, the friction sound effects between the tip of a pen and paper during writing, the sound of turning pages in a book, the ticking of a clock, and even the faint sound of a finger touching the screen are meticulously captured. Some of these crucial sounds are recorded in real-time, while others are selectively sourced from sound effect database.

3.2 Crafting with a Difference - Production Variances in "Immersive" Short Videos

3.2.1 Sound and Image

Upon scrutinizing the outcomes of the comparative study, the distinctions in the production of "immersive" short videos emerge, with a noteworthy emphasis on sound as the predominant feature, while differences in the visual component, although present, are relatively subtle. Hence, the distinctive auditory features constitute the primary audiovisual characteristics of "immersive" short videos.

The auditory dimension in "immersive" short videos holds significant importance, and viewers can understand what is being done by listening it. The specific manifestations include three key aspects. First, the sound in "immersive" short videos prioritizes the original action effects of events, accentuating details in sound, with an emphasis on sound effects such as knocking, tearing, and running water. Sound effects easily overlooked in daily life, such as foam dissolution, lip and tongue touch, and subtle friction, are underscored as essential components of the depicted actions or behaviors. Second, the background sound in "immersive" short videos assumes a distinctive role. The relatively modest use of background music serves to create a specific atmosphere or emotion for the videos—conveying a sense of quiet, soothing, or pleasant ambiance—facilitating viewer immersion into the scene and emotional context. "Immersive" short videos often utilize ambient sound to show the realism of the video,

refraining from sound embellishments and beautification. In some instances, intentionally incorporating noticeable background noise. This unique and sometimes monotonous background sound contributes to a heightened sense of reality during viewing. Examples include the quiet ambiance of doing some collage art in a serene room, barbecuing outdoors accompanied by running water and birdsong, or making beverages in a bustling café. Thirdly, "immersive" short videos tend to minimize voice narration. The use of human voice is also subdued, and even when human voice is present concurrently, it is often delivered in a hushed or breathy manner. Whispering can enhance intimacy, maintaining clarity in explanations and ensuring a certain tone and rhythm to prevent audience monotony. Excessive verbal expression is avoided to prevent disruption of the audience's emotional engagement, allowing them to focus on the depicted events without distraction.

In terms of image, "immersive" short videos exhibit specificity in their approach. When the subject is a person, the focus is on his or her actions and behaviors, with minimal verbal expression from the character. In the case of an object-centric subject, the video emphasizes the object's current state, often employing close-up shots to showcase specific details. Some short videos may even deliberately frame the subject partially outside the screen to provide a more comprehensive presentation of details.

3.2.2 Post-Production Level and the Use of Equipment

The production of "immersive" short videos is characterized by simplicity, utilizing equipment that is accessible to the general public. Consequently, that is the reason why a large number of content creators have emerged, contributing to the widespread proliferation of such videos genre on the internet. Most "immersive" short videos are captured using mobile devices such as smartphones, with some creators opting to employ a separate audio device, usually a portable microphone, to capture key sounds individually.

The production techniques of "immersive" short videos tend to embrace a lifelike audiovisual narrative, favoring minimal shot motion. In comparison to traditional short videos, "immersive" short videos exhibit a more simple filming style, employing fixed shots, automatic or digital zoom, fast shot movement, and other seemingly unprofessional techniques.

Editing "immersive" short videos is less intricate, with a primary focus on sound effects. Some immersive videos remain unedited, presenting raw materials of fixed shot showcasing events from beginning to the end, similar to a live broadcast or raw recording. This simplicity encourages broader public participation in the creation of "immersive" short videos. However, sound post-production in "immersive" short videos involves a special design, emphasizing the importance of key sound elements, whether recorded simultaneously or sourced from existing material in sound effects database.

3.3 Differing Interpretations - Mining the Value of "Immersive" Short Video

3.3.1 The Aesthetic Value of Sound Narrative

The unique aesthetic appeal of "immersive" short videos is attributed to the special aesthetic style infused by sound, contributing to their sustained popularity as a subculture branch within the realm of short videos.

The audiovisual intertextuality plays a crucial role in elevating the audiovisual experience. It refers to the association of music with plot elements, allowing viewers to mentally synchronize the soundtrack with the narrative[6]. In certain food-type "immersive" short videos, quoting specific music or classic film clips heightens emotional resonance and memory recall. For instance, the use of the "Tongue Tip of China" soundtrack in short videos of food production triggers memories and enhances the viewer's immersion in the cooking and eating experience. The memory-evoking capability of music proves invaluable, connecting viewers with past experiences and tastes, fostering a deeper engagement with the video content.

Sound also adds value to images in "immersive" short videos. By highlighting and amplifying sound effects that are easily overlooked. This way of drawing attention to the subtle sound effects, coupled with the very tangible scenes in the image, a synergistic relationship (the marriage of sound and image) is created between the auditory and visual senses. The marriage of the audiovisual, that is, the idea that the two senses, auditory and visual, have a special contrasting and complementary relationship that other senses do not have, and that this relationship is determined by non-technological means. [7] At the level of personal sensation and perception, people get a more complete sense of how sound subtly triggers multiple senses, and therefore, people are able to experience video in

a way that is similar to their daily lives. This audiovisual marriage exploits the unique contrasting and complementary relationship between the auditory and visual senses, contributing to a richer viewing experience.

In terms of the content and narrative of the video, sound narrative fulfills its infinite potential, the effect of which is a win-win situation for both sound and image. We use sound to understand the content of the image, and in turn, the content of the video image contributes to our interpretation of the sound at the same time. Sound narratives expand the available information for viewers, providing nuanced details about the subject's state, material, structure, volume, mood, environment, and more.

People will subconsciously get this additional valuable information through listening after receiving the sound information, which is a process that is not easily perceived by people.

In summary, the aesthetic value of "immersive" short videos primarily stems from the intertextuality of audiovisual culture and the value-added function of sound. The return of auditory senses and the collaboration between the auditory and visual senses alters the way individuals understand media information, emphasizing the importance of sound in shaping the aesthetic experience of "immersive" short videos. According to McLuhan's perspective, this shift from eyes to ears represents a return to auditory space, marking a significant departure from the visual-centric media understanding habits of the past and aligning with the characteristics of a "global village" similar to the tribal era[8]. The synergistic relationship between sound and visuals explains the enduring popularity of "immersive" short videos.

3.3.2 The Commercial Significance of Attraction and Adhesion

The auditory dimension holds significant sway in the realm of "immersive" short videos, as it quickly seizes the audience's attention and promptly stimulates their focus. The centrality of sound in "immersive" short video proves notably advantageous for its dissemination, concurrently brings a certain commercial value. Familiar background music engenders a desire in individuals to persist in their viewing experience. While, the conspicuousness of high volume audio captures attention effectively, fostering curiosity that motivates individuals to continue their viewing experience. Unconventional auditory elements, too, can kindle an exploratory inclination, compelling individuals to sustain their engagement with the video content. The allure of visual effects can captivate individuals, leading to prolonged contemplation, while the appeal of auditory effects initiates a process of curiosity and exploration. The auditory component, acting as an initial stimulus, initiates a sequence wherein visual comprehension lags, prompting individuals to redirect their attention or open their eyes to discern the unfolding events—an inherently explorative process. In this context, sound not only imparts commercial value to "immersive" short videos by instilling a proclivity for ongoing exploration but also aids in the promotion and dissemination of new products.

Moreover, sound in "immersive" short videos cultivates a sense of stickiness, inducing repeated viewing and prompting viewers to actively seek out similar content. This persistence results from individuals developing a reliance on specific auditory elements. A conspicuous example is evident in the comment sections of many Autonomous Sensory Meridian Response (ASMR) videos, where (scrolling) comments such as "sleep aid" "decompression" and "relax" underscore the profound impact of sound. After firsthand observation and analysis, it is evident that the distinctive audio effects in ASMR videos indeed evoke a unique sensory experience. The encounter derived from these particular sound effects stands in stark contrast to that of traditional short videos. Examples include various types of chewing sounds, ear canal friction sounds, and breathy speech sounds, among others, indeed do yield a certain physiological sense of comfort. This experiential sensation dissipates upon the conclusion of the video. In order to prevent the waning of this feeling and to achieve a more enduring state of immersion, viewers tend to develop a viewing stickiness by persistently seeking similar videos or repeatedly watching the same content. . In certain instances, individuals may even experience a desire to make purchases.

In conclusion, the commercial value of sound in "immersive" short videos should not be underestimated. Thoughtful sound designs possess potent allure, capturing attention and facilitating wider dissemination for "immersive" short videos. The auditory component of these videos also engenders viewing stickiness, generating enduring interest that, once established, becomes challenging to dispel. Prolonged and recurrent engagement yields substantial commercial benefits, underscoring the nuanced and impactful role of sound in the "immersive" short video.

3.3.3 The Humanistic Significance of Companionship

The inherent humanistic value embedded in "immersive" short videos should not be underestimated, with the foremost contribution being the companionship they afford, primarily manifesting in a state of virtual coexistence. In contemporary society, the prevalent trend of solitary living is conspicuous, epitomized by the abundance of "empty nesters" not residing in their hometown, often experiencing a pervasive sense of loneliness. Research indicates that loneliness is a negative perception of social isolation, stemming from an inability to integrate into the existing social network [9]. "Immersive" short videos, accompanied by meticulously crafted sound effects, cultivate a state of virtual coexistence. Consequently, the experience of viewing these videos serves as a mitigating force against loneliness, concurrently fostering a sense of belonging within the intricate fabric of the networked society.

Exemplifying this phenomenon are the popular "immersive going back home" short videos on the Xiaohongshu platform, where individuals engage in daily chores upon returning home, such as changing shoes, handling keys, opening windows, washing clothes, and preparing dinner. The showing of whole process going back home in these videos alleviates a sense of loneliness for viewers, imparting a sense of belonging within the virtual community. Observing the daily routines of peers in the video engenders a shared experience, allowing individuals to easily empathize with one another. This immersive sharing of life narratives serves as a unique form of companionship for many.

In summary, sound brings a pivotal humanistic value to immersive short videos, primarily through its companionship function. In an era where societal loneliness has become increasingly pronounced, the need for diverse modes of companionship is paramount. "Immersive" short videos emerge as a source of private assistance, offering psychological comfort and spiritual satisfaction to individuals grappling with heightened feelings of loneliness in the modern life.

4. THE IMMERSIVE EVOLUTION OF SHORT VIDEOS HERALDS A FUTURE OF IMMERSION.

4.1 The Anticipation of Immersive Trends in "Immersive" Short Videos

In the wake of rapid technological advancements, we are gradually entering a future of Immersion. The evolution of "immersive" short videos, representing a distinctive form of media that genuinely immerses its audience, has assumed an increasingly pivotal role in our daily lives. Departing from traditional audiovisual production, it capitalizes on the intrinsic value of sound, drawing viewers closer to a sensory experience reminiscent of virtual reality (VR) and augmented reality (AR) technologies. The development of "immersive" short videos aims to engage audiences within virtual environments through innovative combination of sound and image, foreshadowing a future where immersive development is ubiquitous, bolstered by many more potential advanced technologies.

4.2 The Dichotomy of Impact: Positive and Negative Impacts of Immersive Media

The influence of immersive media on society is double-sided, with coexisting positive and negative implications. On the positive front, immersive short videos offer a novel entertainment form. Leveraging virtual reality technologies, viewers immerse themselves in films, games, music, and other forms of entertainment, heightening their sense of participation and enjoyment. Furthermore, these videos facilitate innovative avenues for learning and training, such as medical students simulating surgical procedures through virtual reality to enhance their understanding and proficiency in surgical operations. Additionally, the collaboration of immersive media with other domains generates richer and more diverse experiences. The integration of immersive media with artificial intelligence technology holds the promise of delivering personalized content recommendations and interactive experiences to users. In the realm of tourism, immersive media serves as a conduit, affording individuals the opportunity to virtually explore global points of interest from the comfort of their homes. Similarly, in the educational sphere, virtual reality augments the learning process by providing students with a more vibrant and intuitive educational experience. As virtual reality technology and augmented reality technology continually advance, individuals will increasingly perceive the virtual world as authentically real. This not only expands cognitive horizons but also engenders novel application scenarios and business opportunities, indicative of the evolving landscape within immersive technologies.

However, we also need to be alert to the possible negative impacts. Firstly, the overindulgence in "immersive" short videos may divert attention from crucial real-world issues, impeding societal progress and problem-solving.

Secondly, excessive immersion in virtual environments may foster detachment from reality, precipitating negative emotional effects such as fear and anxiety. This detachment can lead to increased social isolation, engendering a cycle of addiction to the virtual world, and diminishing real-world communication. Thirdly, the high costs and technical barriers associated with immersive technologies may result in content monopolization and commercialization, limiting creative diversity, and exacerbating the digital divide. Such limitations may marginalize certain groups, impeding their access to the benefits and enjoyment of immersive media, thereby compromising the diversity and inclusivity of society.

In light of these considerations, the application of immersive media, as exemplified by "immersive" short videos, embodies both positive and negative facets. Future endeavors should maximize the positive impacts while remaining vigilant to potential drawbacks. The evolution of immersive media should prioritize diversity and innovation, intertwining seamlessly with other disciplines to offer society richer and more meaningful experiences. We need to simultaneously harness immersive media to its full potential while maintaining attention and engagement with the real world, ensuring that society can achieve comprehensive and sustainable development. Only through a balanced approach to development can immersive media bring about more positive impacts for society and contribute to creating a better future for individuals.

REFERENCES

[1] CNNIC, the 46th "Statistical Report on China's Internet Development", released on September 29, 2020. (http: //www. cac. gov. cn/2020-09/29/c_1602939918747816. htm)

[2] IResearch . China's Short Video Industry Development Research Report 2016[C]//iResearch Series Research Reports (Issue 9, 2016). IResearch; , 2016: 50.

[3] Lin Qianyi. An Investigative Study on the Dissemination of "Immersive" Beauty Short Videos. [J]. Sound Screen World, 2022 (09): 111-113.

[4] Ma Han. Analysis of the Communication Strategy of Tiktok Immersive Short Videos Based on User Psychology [J]. Sound Screen World, 2023 (12): 102-104.

[5] Fu Longyu. Research on the Communication Characteristics of Beauty Immersive Short Videos[D]. Huaibei Normal University, 2023. DOI: 10. 27699/d. cnki. ghbmt. 2022. 000458.

[6] Song Chaoyang. Research on postmodernity and intertextuality of TV drama music in the new media era - taking "Aiqing Gongyu" and "Yanxi Gonglue" as examples[J]. Appreciation of Masterpieces, 2023, (12): 76-78.

[7] Sound: An Acoulogical Treatise. Michel Chion, Duke University Press, 2016

[8] Understanding Media: The Extensions of Man. Herbert Marshall McLuhan, MIT Press, 1995

[9] Loneliness matters: a theoretical and empirical review of consequences and mechanisms. . Hawkley Louise C; ; Cacioppo John T. Annals of behavioral medicine : a publication of the Society of Behavioral Medicine, 2010