ARTICLE

From Ancient Myth to Digital Fires: The Promethean Journey of "Shanghuo" — a TCM Metaphor in Short Video

Shaoci Wang [ID] *, Lay Shi Ng* * [ID] *, Azianura Hani Bt Shaari* [ID]

The Centre for Research in Language and Linguistics, Faculty of Social Sciences and Humanities, The National University of Malaysia, Bandar Baru Bangi, Selangor Darul Ehsan 43600, Malaysia

ABSTRACT

This study, with a focused analysis of "Shanghuo"—a core TCM concept embodying heat pathologies, critically explores the efficacy of digital semiotic systems in illustrating TCM metaphorical concepts. Through a framework integrating principles for effective multimodal metaphor application from Conceptual Metaphor Theory, Multimodal Metaphor Theory, Visual Grammar Theory, and Multimodal Interaction Analysis Theory, this study examines the pattern of multimodal representation of metaphors and performance of four modes—written signs, pictorial signs, sounds, and gestures in translating this metaphorical concept across 20 relevant Douyin videos. The findings reveal that among the current cases of multimodal representation of metaphor of "Shanghuo", strategies combining three or more target modes are more prevalent compared to simple 1-2 modes combinations. However, an examination of their limitations exposes challenges in balancing richness and coherence across different modes. While multimodal integration significantly stimulates viewers' interest, issues like incomplete and contradictory mapping to the target domain—particularly mismatched color schemes in written signs, lack of dynamic design in pictorial signs, and irrelevant audio-gestural cues—not only hinder accurate demonstration of "Shanghuo" but also induce cognitive fatigue in viewers. These findings collectively emphasize the need for optimizing cross-modal consistency and coordination in the multimodal representation of the metaphor of "Shanghuo". Specific recommendations for optimization can be made in two key areas: 1) strengthen the visual, auditory, gestural mapping of fire-associated features to "Shanghuo" in corresponding Zang-organs; 2) enhance cross-modal coordination

and complementarity in mirroring the pathological progression of "heat" evolving from the five Zang-organs to the body's surface.

Keywords: Multimodal Representation of Metaphor; Short Videos; Traditional Chinese Medicine; "Shanghuo"

1. Introduction

1.1. Research Background

During the COVID-19 pandemic, TCM gained prominence in China for its role in prevention and treatment strategies[1, 2]. With social media emerging as a key medium for information dissemination, platforms like Douyin hosted a surge of TCM short videos content, covering topics from TCM interpretation of pathology to TCM dietary and treatment advice. To engage a broader audience, even videos shorter than one minute incorporate subtitles, voiceovers, pictorial signs, music, and sound effects[3–5]. Such short videos won significant popularity due to their use of diverse multimodal strategies[6–9]. While multimodal representation greatly enhances user engagement, there are widespread concerns that the oversimplified, entertainment-driven multimodal designs will dilute TCM's scientific rigou[10, 11]. This study therefore aims to conduct a case study on the multimodal representation of "Shanghuo", an important TCM metaphorical concept, within short videos on Douyin platform. The objective is to assess the effectiveness of multimodal strategies in conveying TCM metaphorical concepts via short videos and to propose corresponding improvement measures, addressing the prevalent challenge of balancing accessibility and credibility of digital discourse in disseminating TCM knowledge.

1.2. Problem Statement

TCM popularization short videos of various kinds garnered wide academic attention in China in the post-epidemic era[12–14]. Existing studies consistently highlight that the multimodal integration of text, imagery, sound, and physical representation in TCM popularization short videos reduces the abstraction of TCM knowledge and enhances audiences engagement[2, 15]. By translating TCM terminologies (e.g. herbal medicines or specific acpoints) into accessible audiovisual formats, these videos effectively bridge the gap between specialized medical concepts and public understanding[1, 6]. However, many scholars widely show a concern: the entertainment and accessibility brought by multimodal strategies have been diluting the scientific rigor of TCM knowledge, hindering the accurate representation of TCM's theoretical depth and dialectical thought[8, 10, 11].

Such a negative effect of the multimodal strategies is assumed to stem from a failure to design multimodal representation grounded on the metaphorical logic inherent to the TCM system. It is found that the TCM system consists of complicated TCM concepts and philosophical theories that are extensively built on conceptual metaphors. The understanding of human health and bodily functions within the TCM system is deeply intertwined with nature and philosophical thought. It is based on the perceptions of the world and the understanding of the traditional Chinese philosophy that ancient Chinese practitioners utilized a conceptual metaphor system to develop the complex TCM frameworks that explain bodily functions and pathology[16–24]. As stated in *Huangdi Neijing (The Yellow Emperor's Classic of Medicine)*, "Failure to use analogy leads to unclear knowledge"[25]. The method of analogy and imagery, as a primary construct tool in TCM system, has played a significant role in the formation of TCM concepts and theoretical systems. Conceptual metaphors are found pervasive throughout the TCM system[17].

Multimodal strategies risk failing to harness their full potential in demonstrating scientific and cultural value of TCM knowledge through short video if multimodal elements are superficially superimposed as mere attention-grabbing gimmicks, rather than being systematically mapped onto the conceptual metaphor system of TCM. This disconnect becomes critically apparent when explaining core TCM concepts like "Shanghuo" that inherently rely on conceptual metaphors. "Shanghuo" (上火), a key concept related to febrile syndromes in TCM theory, primarily refers to exuberant internal heat caused by the imbalance of Yin and Yang in the body. It can be embodied in the five Zang-organs, including heart heat, lung heat, liver heat, kidney heat, and spleen heat. "Shanghuo" will be manifested by a series of

fire-associated traits like "burning", "redness and swelling", "rising motion", "consumption", "agitation", "dryness or cracking", etc. While this concept is commonly depicted through the universal metaphor of "fire" to signify inflammation in human body, its clinical complexity lies in the distinct symptoms stemming from internal heat in different organs. According to Zang-Xiang Theory in TCM system, the heart opens its orifices in the mouth; the lungs open their orifices in the nose; the kidneys open their orifices in the ears; the spleen opens its orifices in the teeth; the liver opens its orifices in the eyes[25]. If the heart suffers from the excess heat, it will be manifested by "redness in the mouth"; if the lung suffers from the excess heat, it will be manifested by "swelling and redness in the nose"; if the liver suffers from the excess heat, it will be manifested by "dryness in the eyes and irritation"; if the kidney suffers from the excess heat, it will be manifested by "roaring in the ear"; if the spleen suffers from the excess heat, it will be manifested by "swelling and heat in the gum". Therefore, if the concept of "Shanghuo" is simply presented by generic symbols—"red flames" without employing specific multimodal strategies to map the corresponding syndromes features, it will risk superficial or even misleading medical diagnoses and treatment suggestions, ultimately leading to prioritizing entertainment over scientific rigor, and diminish trust in TCM among audiences.

Based on the identified problem, this study attempts to focus on Douyin short videos illustrating "Shanghuo" embodied in five Zang-organs, to examine how the current multimodal strategies translated this mystical metaphor. Specifically, this study aims to 1) investigate the multimodal co-occurrence patterns and performance of each mode in representing the related concepts; 2) expose the limitations of these multimodal representation and how they hindered the accurate and intuitive illustration of this concept; 3) propose practical improvements in the multimodal strategies, thereby enhancing the clarity of metaphorical logic of multimodal representation. With the increasing popularity of TCM popularization short videos, it is crucial to improve the fidelity and objectivity of the multimodal presentations of TCM concepts. Such improvements are greatly significant for not only enhancing public understanding but also advancing the global dissemination of TCM as both a medical practice and a cultural heritage[8, 15].

1.3. Theoretical Framework

This study will combine Conceptual Metaphor Theory[26], Multimodal Metaphor Theory[27], Visual Grammar Theory[28], and Multimodal Interaction Analysis Theory to build a theoretical framework to explore the effectiveness of the multimodal representation of metaphor[29]. These theories together provide insight into the principles for the effective multimodal metaphors. On the basis of this, this research proposes hypotheses and makes an analysis of the collected data.

Conceptual metaphors involve mapping one conceptual domain (the source domain) onto another (the target domain), allowing individuals to comprehend intangible concepts through more concrete experiences. As the core principle of Conceptual Metaphor Theory, the systematic mapping principle emphasizes the consistency of cognitive topology and logical structures of the source and target domains. This principle posits conceptual metaphor shall ensure inferential coherence through holistic rather than isolated element mappings; otherwise, it leads to cognitive dissonance and communication breakdown[26]. Multimodal Metaphor Theory proposes that non-verbal elements, such as images, sounds, spatial layouts, and colors, can also be involved in constructing metaphors to convey concepts. This study adopts the broad definition of multimodal metaphor, wherein mapping from the source domain onto another the target domain could be completed through a single dominant mode, with other modes acting as amplifiers[27, 30]. Expanding on the Conceptual Metaphor Theory, Multimodal Metaphor Theory[27], Visual Grammar Theory [28], and Multimodal Interaction Analysis Theory together develop the principles for the effective application of multimodal metaphors[29]. Multimodal mapping, like conceptual metaphor, is suggested to preserve the same inferential structure of the source domain as the target domain to maintain epistemological fidelity, avoiding oversimplification or symbolic mismatches that erode conceptual nuance[27]. In addition, a clear multimodal metaphor is featured by cross-modal consistency—aligning visuals, sounds, and text to reinforce a unified metaphorical logic, ensuring logical congruence through symbols resonant with the audience's cognitive schemas[27]. Moreover, the effective representation of multimodal metaphors relies on strengthening the modal coordination in complementarity and tem-

poral adaptation[29]. What's more, the design of multimodal metaphor should prioritize perceptual or cognitive salience across modes and allocate properly the modal weighting according to the focus of metaphor[28, 30]. Finally, the adoption of dynamic adaptation strategy helps multimodal representation retain the narrative coherence according to the specific logic of metaphor[27, 29].

These principles enable this study to make hypotheses and a comprehensive examination of the effectiveness and limitations of multimodal strategies. This study focuses on 20 five-Zang-organ -related "Shanghuo" short videos on Douyin platform, aiming to 1) identify the multimodal representations of metaphors regarding five-Zang-organ-related "Shanghuo", 2) assess their limitations in illustrating the metaphor of "Shanghuo", and (3) offer recommendations for enhancing the accessibility and accuracy of such short videos.

2. Materials and Methods

2.1. Data Collection

This research established datasets of multimodal metaphor of five-Zang-organ-related "Shanghuo" collected from 20 relevant Douyin short videos. The type of modes presented in selected short videos were classified according to Forceville's classification: pictorial signs, written signs, spoken signs, gestures, sounds, music, smells, tastes, and touch[30]. This study focuses on four of them—written signs, pictorial signs, sounds, and gestures, and defaults voiceover (spoken signs) as the anchoring mode. The short videos are selected based on the following criteria: 1) the videos are characterized by distinct multimodal designs involving at least two target modes to illustrate the concept of "Shanghuo"; 2) videos are limited to a maximum of three minutes in duration and have garnered a minimum of 1,000 likes to ensure analytical feasibility and social relevance; 3) each video must feature real practitioners on camera, accompanied by voiceover narration for contextual clarity.

2.2. Research Methods

This research will apply a qualitative method to conduct the analysis. Specifically speaking, it will be carried out in four stages as follows:

The first stage is to transcribe the selected short videos to sort out datasets of metaphors in each selected short video through Descript (an online transcription software). Upload the selected TCM short videos to Descript and build a project for each short video. Using the Transcript function to transcript the voiceover of the selected short videos. To import the scripts and to build datasets of metaphors.

The second stage is to design hypotheses regarding the limitations of multimodal representation for metaphors based on the above-mentioned principles as follows:(1) the modes representing the source domain and that representing the target domain are inconsistent; (2) the modes representing the source domain are mismatched or contradictory; (3) the weight of modes representing the source domain is not allocated improperly; (4) the modes representing the source domain are lack of dynamic narrative; (5) the complementarity of modes representing the source domain is weak.

The third stage is to collect and process data through Eudico Linguistic Annotator 6.9 (ELAN) and Microsoft Excel 2021.

1) To formulate the coding scheme based on the above hypotheses. The coding scheme divides two parts: performance of modes, and distribution of modes usage in a unit of metaphor. The source domain and the target domain are coded as [S] and [T], respectively. Written signs, pictorial signs, gestures and sounds are coded as [w], [p], [g] and [s], respectively. In this study, the voiceover is defaulted as the anchor of metaphors, thus [vt] means voiceover representing the target domain and [vs] means voiceover representing the source domain.

As for performance of each mode, if the mode representing the source domain is inconsistent with (vaguely associated with or totally conflicting with) the mode representing the target domain, it will be coded as [SNC]; if the mode representing the source domain is lack of dynamic design, it will be coded as [SND]; if mode representing the source domain appears too fast or imperceptibly, it will be coded as [SNS]; if the mode representing the source domain is correct and clear, it will be coded as [SQ]; [SA] is used to show the mode representing the source domain is absent.

As for distribution of modes usage in a unit of metaphor, [MC] refers to that the modes representing the source domain are conflicting; [MI] refers to that the weighting of modes representing the source domain is allocated improp-

erly (where the modes with low perceptual salience exhibited dominance over that with high perceptual salience); [MOS] refers to that the modes representing the source domain are out of sync; [MCW] refers to that the complementarity of modes representing the source domain is weak (the modes representing the source domain are simply repeated or one of them exhibits irrelevant mapping); [MSP] refers to the co-occurrence pattern of the modes representing the source domain.

The coding scheme is shown in **Table 1**.

Table 1. The Coding Scheme.

Code	Meaning
[T]	the target domain
[S]	the source domain
[vt]	voiceover (the target domain)
[vs]	voiceover (the source domain)
[w]	written signs
[p]	pictorial signs
[s]	sounds
[g]	gestures
[SNC]	the mode representing the source domain is inconsistent with (vaguely associated with or totally conflicting with) the target domain
[SNS]	the modes representing the source domain appears too fast or imperceptibly
[SND]	the mode representing the source domain is lack of dynamic design
[SQ]	the mode representing the source domain is correct and clear
[SA]	the mode representing the source domain is absent
[MC]	the modes representing the source domain are contradictory
[MI]	the weighting of modes representing the source domain is not allocated improperly (the modes with low perceptual salience are over-dominance)
[MOS]	the modes representing the source domain are out of sync
[MCW]	the complementarity of modes representing the source domain is weak (the modes representing the source domain are redundant, replicated, or one of them exhibits irrelevant mapping)
[MSP]	the co-occurrence pattern of the modes representing the source domain

2) To annotate the data in ELAN 6.9. ELAN 6.9 enables playback of short videos in a loop with precise positioning to 0.1 seconds, allowing to identify the multimodal representation of the source domains and the target domains. This study will focus on 4 modes including pictorial signs, written signs, gestures, and sounds engaged in constructing multimodal metaphors in selected short videos. Upload the short videos in ELAN 6.9 and create 2 primary annotation tiers: performance of modes and distribution of modes usage in a unit of metaphor. Under the tier of performance of modes, to build 6 secondary tiers naming them by vt, vs, p, w, g, and s representing the targeted modes. The performances of modes representing the source domains and the target domains will be annotated in these tiers respectively according to the coding scheme mentioned above.

Under the tier of distribution of modes usage, to build 5 secondary tiers naming them by MC, MI, MOS, MCW, MSP. The modes of usage will be annotated respectively in these tiers according to the coding scheme. To export the datasets of multimodal representation of "Shanghuo" after annotation.

The specific annotation instructions are as follows.

If written sign is contradictory with voiceover, it will be coded as [vs-w] in the tier of MC, and so on;

If non-visual signs (e.g., sounds) exhibit over-dominance, they will be coded as [s>] in the tier of MI;

If written sign is not synchronized with voiceover, it will be coded as [vs\w] in the tier of MOS;

If the modes representing the source domain are simply duplicated, [md] will be coded in the tier of MCW; if one of them exhibits irrelevant mapping, [mi] will be coded in the same tier;

If voiceover, written sign, picture, and sounds are used to represent the source domain, [vwps] will be annotated in the tier of MSP to show the the co-occurrence pattern, and so on.

3) To make a further data process in Microsoft Excel 2021. Excel allows the generation of interaction network maps or data visualizations, such as frequency analysis. First, to identify the frequency of co-occurrence patterns of different modes representing the source domain and target

domain consisting of multimodal metaphors. Second, to analyze the distribution of limitations of the use of multimodal metaphors.

The final stage is to discuss the results based on the collected data and draw conclusions. To summarize the multimodal co-occurrence patterns of metaphors used in the selected TCM short videos, and analyze the limitations of the application of these multimodal metaphors. Based on that, it will offer recommendations for improving the effectiveness of the multimodal metaphors usage in conveying the concept "Shanghuo".

3. Results

The total 151 datasets of multimodal metaphor have been sorted out from the 20 selected videos. The detailed analysis of these datasets reveals co-occurrence patterns, performance attributes, and distribution of target modes.

3.1. Modal Co-Occurrence Patterns

The analysis of modal co-occurrence patterns indicated a significant preference for multimodal complexity in the representation of metaphors. Notably, 25.17% of the metaphors incorporated a combination of voiceover, written signs, pictorial signs, sounds, and gestures ([vwpsg]), whereas less than 5% of cases used simpler combinations, such as those including only voiceover and written signs ([vw]). Intermediate patterns disparities across combinations. Patterns consisting of voiceover written signs and sounds ([vws]) represent 19.21% of the total, as well as the combination of voiceover, written signs, pictorial signs, and sounds ([vwps]) constitutes 13.91%. In contrast, those including voiceover paired with written signs and gestures or pictorial signs([vwg]/[vwp]), together account for approximately 10%. These findings underscored the dominant roles played by written signs in the construction of multimodal metaphors within the analyzed videos (see **Figure 1**).
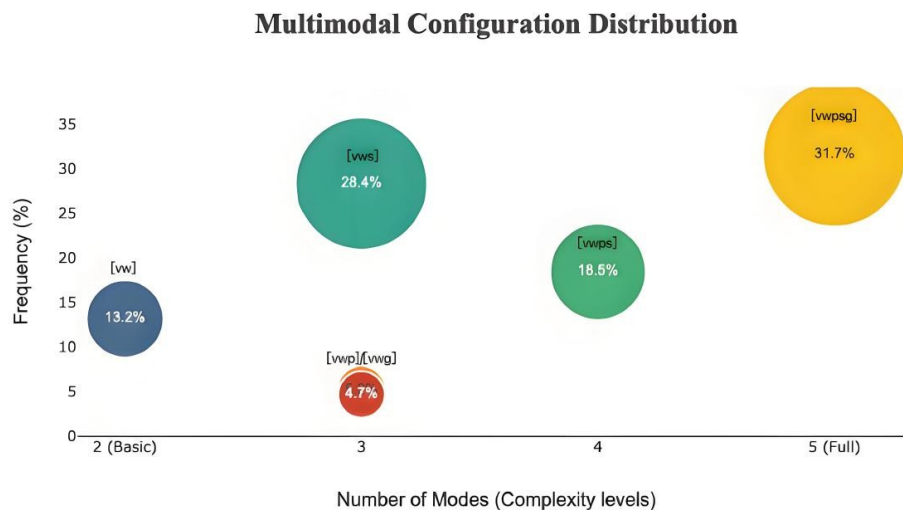

Figure 1. Multimodal Configuration Distribution.

3.2. Modal Performance and Distribution Analysis

Modal performance analysis revealed significant disparities in the consistency and clarity across the different modes utilized. As a constant visual element in all types of combinations, written signs demonstrated the strongest performance, with 67 cases classified as correct and clear (SQ), representing 44.3% of the total. However, 37.1% of written signs were inconsistent with target domains ([SNC]), indicating frequent mismatches between textual and vocal narratives. Despite their overall good performance, around 18.5% of written signs were noted to appear too briefly or imperceptibly (SNS). Conversely, pictorial signs—as another critical visual element —were absent in 44.8% of cases([SA]). Nevertheless, pictorial signs demonstrated the lowest level of inconsistency of mapping ([SNC], 23.8%). It is noteworthy that a significant 32.4% of cases faced substantial challenges

in dynamic design([SND]), where static or underdeveloped imagery failed to reinforce metaphorical intent. This resulted in only 17.2% being categorized as correct and clear ([SQ]).

Both sounds and gestures demonstrated pronounced inconsistency, with sounds exhibiting a 47% inconsistency rate ([SNC]) and gestures at 27.8%. Furthermore, 15.8% of sounds were found to be unclear or disappeared too quickly for viewers to process (SNS), with only 5.2% being classified as clear (SQ). Gestures faced similar challenges, with 9.9% classified as having insufficient duration or being imperceptible ([SNS]), which diminished their communicative impact. Only 6.6% of gestures were categorized as clear (SQ), further highlighting the confusing role that gestures played in the multimodal representation of metaphors (see **Figure 2**).

Modal distribution analysis exposed other correlations influencing mapping effectiveness. It signified that all modes were involved in varying degrees of modal conflicts ([MC]), with 33.1% of such conflicts involving written signs, 23% sounds, and 15.8% gestures, 4.6% pictorial signs. Notably, the rate of inconsistency between written signs and voiceover nearly coincided with the proportion of written signs involved in modal conflicts, underscoring that the core challenge of written signs lay in the contradictory mapping to the target domain. In contrast, though the inconsistency between sounds and voiceover was most prominent([SNC], 47%), the involvement of sounds in modal conflicts was not the highest, indicating that the main comprehension barriers posed by sounds resulted from their vague mapping to the target domain. (see **Figure 3**).
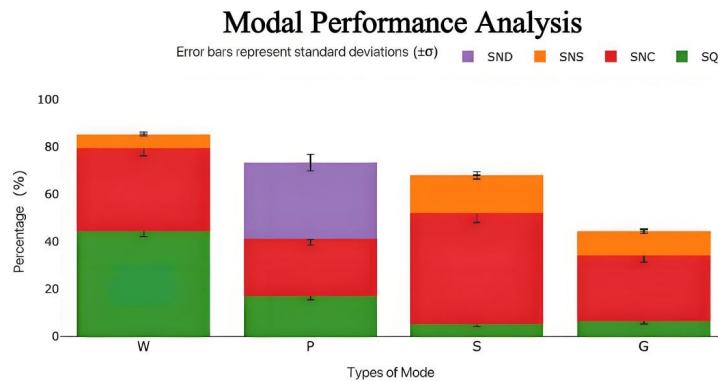

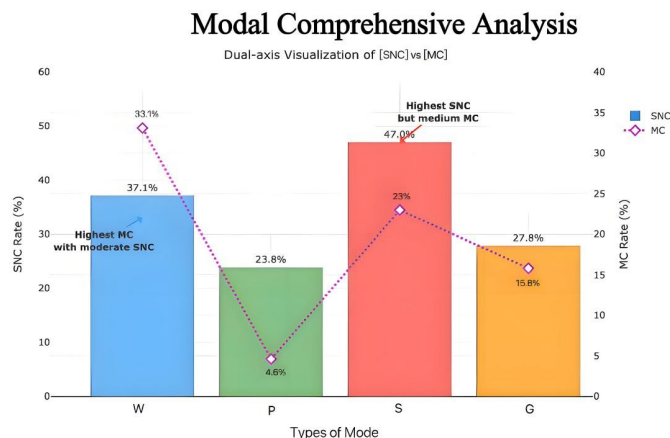**Figure 2.** Modal Performance Analysis.

Figure 3. Modal Comprehensive Analysis.

In addition, although the dominant role of visual signs in all cases resulted that the issue of improper weight allocation of modes representing the source domain was almost nonexistent ([MI], 0%), the high prevalence of absent pictorial signs([SA], 44.8%) showcased the risk of over-reliance on written signs and neglect of visual supplementation from pictorial signs, indicating that the monotony in application of visual elements in the multimodal mapping strategies (see **Figure 4**).

Moreover, despite the rare synchronization issues ([MOS], 3.3%), as high as 64.8% of total cases exhibited weak complementarity in cross-modal mapping ([MCW]).

Among these, the problem of cross-modal redundant duplication accounted for 33% of total cases, whereas, in 47% of total, the weak coordination was manifested by the modal irrelevant mapping. A further correlation emerged when comparing the distribution of these two issues across different modal co-occurrence patterns: the pattern integrating all target modes ([vwpsg]) represented the highest proportion of cross-modal redundant duplication (30 out of 48 cases), while the pattern combining voiceover, written signs with sounds([vws]) faced more severe challenges in modal irrelevant mapping than other patterns (30 out of 43 cases) (see **Figure 5**).
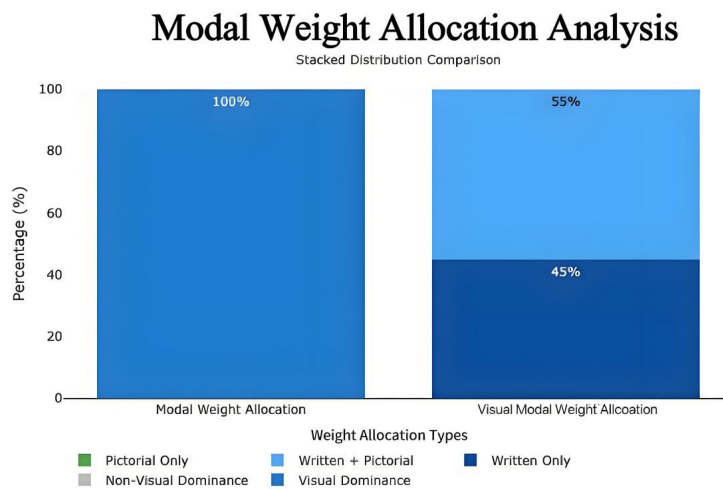

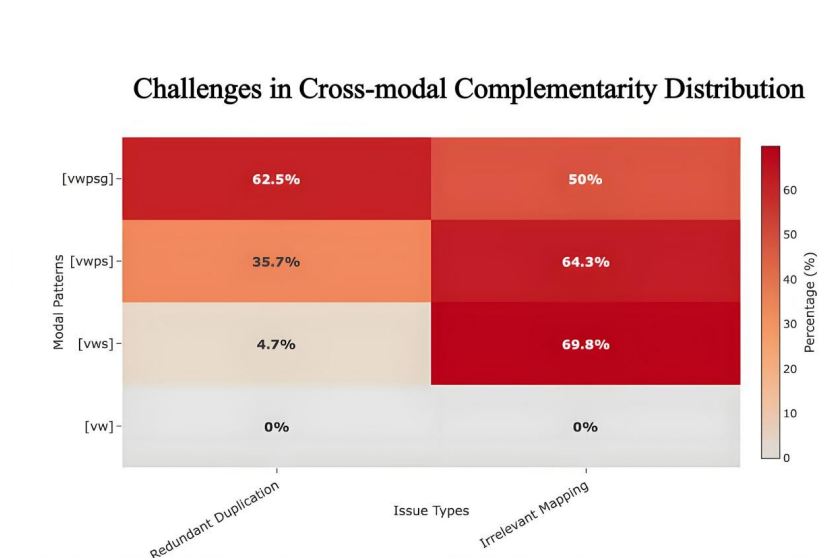**Figure 4.** Weight Allocation Analysis.

Figure 5. Challenges in Cross-modal Complementarity Distribution.

Overall, the data analysis highlighted the tension between richness and coherence in the multimodal design for the metaphor of "Shanghuo". These findings emphasize the need to address identified limitations in multimodal representation, ultimately enhancing the effectiveness of multimodal metaphors usage in conveying TCM metaphorical concepts. Subsequent sections will present further analysis and recommendations to support these improvements.

4. Discussion

Base on the aforesaid data analysis, this part will discuss the result and make some recommendations to enhance the effectiveness of multimodel representation of metaphorical concept of "Shanghuo".

4.1. Limitations of the Application of Written Signs and Recommendations

In current practices, approximately half of written signs performed satisfied in the multimodel representation of metaphorical concept "Shanghuo". The content of these cases primarily aligned completely with the voiceover, which served as anchors in this study. In addition, the colors and forms used were carefully chosen to ensure visual consistency with characteristics of "fire". Specifically, these designs predominantly employed red, orange, or yellow hues for written signs, reinforcing associations with symptoms like "heat", "redness", "swelling", etc. What's more, a subset of these designs did better by incorporating dynamic typography. For instance, flickering text animations were used to illustrate "liver opens its orifices in the eyes, thus excess liver fire is manifested by dry eyes and irritability", effectively mirroring the quality of "agitation" associated with fire. Such approaches not only maintain consistency across modes representing the source domain, but reduce the abstractness of pure text, helping audiences grasp features of syndromes in an intuitive way.

However, around one-third of analyzed videos revealed various types of inconsistencies in mapping. Typical issues included mismatched color schemes, such as white or blue fonts, which failed to evoke the heat or intensity of fire—blue, in particular, risks implying "coolness", leading to a contradictory mapping onto the target domain (see **Figure 6**). Another issue lay in the indiscriminate use of decora-

tive font templates. Taking the cracked-font as an example, while it effectively symbolized "dry lips caused by heart heat", their application in mapping onto unrelated symptoms, such as "nasal redness from lung heat", created confusion by conflating distinct attributes. Additionally, overly dense text layouts and fleeting on-screen duration were assumed to overwhelm viewers, exacerbating cognitive fatigue.

Figure 6. White Fonts Representing "Shanghuo" Failed to Evoke the Features of Fire.

To optimize these multimodal strategies, firstly, color discipline remains critical. The choice of color should strictly adhere to warm hues (red, orange, yellow) to preserve metaphorical coherence. Moreover, a tailored typography is essential. Font design is recommended to align closely with corresponding symptoms; for instance, designing red, rounded fonts to map onto the syndrome like "red and swelling" while utilizing vibrating text to imply symptoms like tinnitus or irritability. Last, the density of the text displayed should be adjusted. The flexible combination of text and images is suggested to replace the demonstration of text repeating the voiceover, and leave the task of mapping the features of symptoms to the pictorial signs. For instance, as for the multimodal representation of the metaphor "the spleen opens its orifices in the teeth, excess spleen fire causes swollen and painful gums", red text could continuously display "the spleen opens its orifices in the teeth" until the metaphor ends. At the same time, corresponding pictorial signs, such as animations of flames burning in the mouth, could be used to reinforce dynamic mapping. This strat-

egy would not only ensure consistency of mapping from the source domain onto the target one, but also avoid information redundancy and alleviate cognitive load on the audiences.

4.2. Limitations of the Application of Pictorial Signs and Recommendations

The data analysis revealed that critical challenges in the utilization of pictorial signs mainly stemmed from inconsistent and incomplete mappings, static visual design, and mode redundancy.

First, incomplete or inconsistent mappings caused by pictorial signs significantly undermined the clarity and even caused confusion of multimodal representation. For instance, a static ear imagery was used in multimodal representation of the metaphor "kidney heat causes tinnitus", where the representation of fire-associated traits (e.g., the sound of soaring) was omitted. This fragmented mapping left audiences unable to intuitively perceive the symptoms of kidney heat. Another notable issue was the undifferentiated representation of "Shi Huo" (实火,excess internal heat) and "Xu Huo" (虚火,fluid-deficiency induced heat). A typical example can be found in the animations depicting "Yin deficiency-induced kidney heat" through generic flame imagery (see **Figure 7**), which directly contradicted the logic of TCM pathology: "Xu Huo" originates not from excess internal heat but from fluid depletion in human body[25]. To improve this, it is suggested to visualize a cross-section of a kidney displaying low water levels along with smoldering embers. This approach would be more effectively showcase the distinct physiological and pathological mechanisms underlying different types of "Shanghuo".

Second, static imagery limited the potential of multimodal strategies to depict the pathological development of "Shanghuo". A typical example can be revealed in the adoption of static visuals of a kidney and an ear surrounded by wavy lines in mapping the concept that "kidney heat is manifested by tinnitus". The viewers will hardly understand how the ear-related symptoms were connected with the kidney heat from two static pictures if the pathological development of a disease is not illustrated. The adoption of temporal adaptation in the design of pictorial signs is conducive to addressing the problem by illustrating the relation of symptoms at different bodily parts. Compared with the static images of a kidney and an ear with wavy lines, designing an animation of a cross-section of a kidney with low water levels along with smoldering embers paired with an animated ear that vibrates in conjunction with a fire-suggestive "roaring" sound is more likely to clarify the chain reaction from the kidney heat to a painful ear. Therefore, the application of dynamic strategy is particularly critical for the intuitive mapping onto the pathological progression of "Shanghuo" embodied in five Zang-organs.

Figure 7. Animations Depicting "Yin Deficiency-Induced Kidney Heat" Through Generic Flame Imagery Directly Contradicted the Logic of TCM Pathology.

Third, the redundancy of pictorial signs reflected weak cross-modal complementarity, which not only reduced the coherence of multimodal metaphor but exacerbated cognitive overload. A striking example of information duplication appears in videos explaining "heart heat is manifested by mouth ulcers", where ulcers-decorative text, image of mouth with ulcers, and a doctor's gesture of pointing to the mouth were displayed simultaneously. An optimized strategy would replace the static image of mouth ulcers with a simple gesture of pointing to the mouth, triggering context-specific animations, like small animated flames showing on the lips of doctor, pared with auditory cues, like ire-like hissing sounds. This multimodal integration strategy, which assigns representational roles based on the unique strengths of each mode, will helps systematically minimizes redundant encoding and crucially reduces the visual noise.

4.3. Limitations of the Application of Sounds and Gestures and Recommendations

The analysis of sounds and gestures performance in multimodal strategies showed significant limitations in mapping consistency and cognitive load management. Sound emerged as the mode with the highest proportion of inconsistent mappings, primarily due to the mechanical use of entertainment-oriented audio effects offered by the platform. Approximately half of the problematic sounds incorporated generic sound cues (e.g., repetitive "dings" or "clicks") to signal new content, a strategy initially intended to heighten viewer alertness but which instead caused cognitive fatigue and fragmented information processing. In other circumstances, sounds effect was suggestive of unrelated symptoms. For example, a fading whirling sound simulating dizziness was used to illustrate the metaphor "heart heat is manifested by headaches", which did not strictly align with the syndrome of headaches induced by heart heat. Such inconsistencies from misused sounds led to misleading mappings and weakened metaphorical coherence. Another key issue was the imperceptibility of sounds due to low volume or insufficient duration, which prevented sounds from fulfilling their potential in scenario-specific mapping.

It is suggested to improve sound usage by reducing the frequency of unrelated sound effects to prevent distracting the audience from key visual information. Additionally, replacing generic audio by the sounds effect suggesting flames burning, such as "sizzle", "beep-boop", or "roaring", to assist in symptom-specific mapping (e.g., using "crackling" for mild "spleen heat" vs. "roaring" for severe "heart heat"). Finally, ensuring the salience of these sound effects within scenario-specific multimodal contexts will optimize their effectiveness in reinforcing metaphorical coherence.

Gestures also faced a challenge in incomplete mapping due to physicians' habitual or redundant hand movements. Many on-camera doctors' frequent employment of arbitrary hand motions aligned with speech rhythms offered very limited mapping cues, and easily distracted viewers from essential visuals. Moreover, insufficient amplitude or duration limited the effectiveness of intentional gestures. In the scenario where an undersized or fleeting gesture of pointing to the ear for the mapping of metaphor "kidney heat is manifested by tinnitus", viewers often missed the transient cues of gestures before processing their relationship to accompanying text or images, which led to the fragmented cross-modal mapping (see **Figure 8**).

Figure 8. Presenter's Undersized or Fleeting Gesture of Pointing to the Ear Led to Insufficient Gestural Mapping.

To optimize gesture use, it is essential to strengthen the coordination of gestures with other visuals. Doctors on-camera can be trained to point precisely to the affected areas with their hands and maintain the gesture until the multimodal mapping is complete. For example, to visualize the concept of "kidney heat is manifested by tinnitus", an animation could depict a cross-sectional view of the kidney with diminished water levels and smoldering embers, accompanied by an animated ear vibrating to "roaring" sound effects evoking a fiery sensation. Concurrently, the on-camera doctor could enhance the tinnitus mapping by gesturing with a trembling palm around the animated ear, enabling viewers to intuitively grasp the causal chain from kidney heat to ear pain. During other segments, they should keep their hands still to minimize information redundancy and visual clutter.

5. Conclusions

In conclusion, the analysis of modal co-occurrence pattern reveals that the current cases of multimodal representation of the metaphorical concept "Shanghuo" give remarkable priority to the adoption of multimodal combinations over simpler modal combinations. However, after examining the limitations in the performance and distribution of the four target modes, several key findings emerge. First, the

hypothesis that the four target modes (as source domain representations) exhibit varying degrees of inconsistency with the target domain is validated. Second, while the assumption of improper allocation of modal weights in these multimodal representations is not confirmed, the results reveal an imbalance in the application of visual modes within multimodal representations. Third, the prevalence of the problem of cross-modal redundant replication of information and the issue of modal irrelevant mapping, attest to the hypothesis of weak cross-modal coordination. Finally, the failure to map the pathological progression of "Shanghuo" demonstrates that the lack of dynamic narrative significantly limits the potential of pictorial signs as a critical visual element in illustrating TCM metaphorical concepts.

These findings collectively underscore the importance of optimizing multimodal metaphors for the concept of "Shanghuo". Specific improvements can be made in two key areas: 1) strengthen the visual, auditory, gestural mapping of fire-associated features to "Shanghuo" in corresponding Zang-organs; 2) enhance cross-modal coordination and complementarity in mirroring the pathological progression of "heat" evolving from the five Zang-organs to the body's surface. Such improvements are anticipated not only to alleviate viewers' cognitive fatigue effectively but also to visually articulate the mechanism of the inflammatory processes, thereby addressing the existing imbalance between the accessibility and accuracy of the related short videos.

This study critically examines the limitations of digital semiotic systems in interpreting TCM metaphorical concepts from the perspective of the effective use of multimodal metaphor. Diverging from prior research that primarily highlighted the role of digital discourse in popularizing TCM knowledge, this work unveils how the indiscriminate application of generic multimodal symbols (e.g., universal "fire" imagery) dilutes the intrinsic logic of TCM's metaphorical system. It further proposes actionable strategies to optimize the multimodal representation of metaphorical concepts in TCM short videos. While offering novel insights into the digital representation of "Shanghuo" metaphors, this study has several constraints. First, the limited sample size may affect the generalizability of the research results. Second, the coding of multimodal metaphors, despite rigorous protocol adherence, risks researcher bias in interpreting modal performance. Therefore, future research can further strengthen

objectivity by exploring the following directions: for example, using inter-rater reliability assessments to objectify semantic annotation; or verifying through experiments such as eye-tracking tests to quantify how multimodal presentations of metaphors affect viewers' understanding of TCM metaphorical concepts.

Metaphorical systems form the cognitive foundation for comprehending TCM knowledge[17]. When conveying TCM metaphors through digital semiotic systems, reducing them to mere rhetorical decorations risks undermining both the credibility of TCM and the integrity of its theoretical framework. By strategically leveraging the capability of multimodal strategies in creating an intuitional experience, it becomes possible to align aesthetic engagement with scientific rigor. Such a semiotic framework, rooted in TCM's inherent metaphorical logic, can transform the viral communication route into a conduit for authentic knowledge transmission. This approach ensures that digital platforms evolve beyond superficial entertainment, instead serving as bridges that connect ancient TCM wisdom to contemporary health literacy, thereby influencing public health awareness and behaviors.

Author Contributions

Conceptualization, S.W. and L.S.N.; methodology, S.W.; software, S.W.; validation, L.S.N. and A.H.B.S.; formal analysis, S.W.; investigation, S.W.; resources, S.W. and L.S.N.; data curation, S.W., L.S.N. and A.H.B.S.; writing—original draft preparation, S.W.; writing—review and editing, S.W., L.S.N. and A.H.B.S.; visualization, S.W.; supervision, L.S.N. and A.H.B.S. All authors have read and agreed to the published version of the manuscript.

Funding

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Acknowledgments

Confict of Interest

The authors declare no conflict of interest.

References

[1] Su, R., 2022. Research on the Visual Discourse of Health Science Short Videos [Master's Thesis, in Chinese]. Heilongjiang University: Harbin, China. DOI: https://doi.org/10.27123/d.cnki.ghlju.2022.000328

[2] Wang, Q., Gu, Y., Xiao, Q., et al., 2021. Multimodal Discourse Meaning Construction of TCM Culture Media [in Chinese]. Science Communication [in Chinese]. 7, 72 – 76.

[3] Feng, F., 2023. An Analysis of the Application of Humor in Health Popularization Short Videos by "Dingxiang Doctor" on the Douyin App [Master's Thesis, in Chinese]. Guangxi University: Nanning, China. DOI: https://doi.org/10.27034/d.cnki.ggxiu.2023.000096

[4] Li, X., 2023. Research on the Design of Health Science Videos in the New Media Era: A Case Analysis of the Content of the Science Self-media "Mr. Sixth Floor" [in Chinese]. Science Communication [in Chinese]. 19, 30 – 34.

[5] Zhang, C., 2023. Visual Practices, Practical Problems, and Solutions in Health Popularization Short Videos: A Multimodal Perspective [in Chinese]. Chinese Culture and Communication Research [in Chinese]. 2, 193 – 201.

[6] Chen, Z., 2023. A Study on the Dissemination of TCM Health Short Videos on the Douyin App [Master's Thesis, in Chinese]. Hunan University: Changsha, China. DOI: https://doi.org/10.27135/d.cnki.ghudu.2023.002193

[7] Cheng, B., 2022. A Study on the Content Production of TCM Short Videos on the Douyin Platform [Master's Thesis, in Chinese]. Hainan Normal University: Haikou, China. DOI: https://doi.org/10.27719/d.cnki.ghnsf.2022.000013

[8] Jing, Q., 2023. Challenges and Breakthroughs in the Dissemination of TCM Culture Short Videos [Master's Thesis, in Chinese]. Shandong Normal University: Jinan, China.

[9] Wang, H., 2022. A Study on Content Production of Health Communication Short Videos by Doctor KOLs [Master's Thesis, in Chinese]. Tianjin Normal University: Tianjin, China. DOI: https://doi.org/10.27363/d.cnki.gtsfu.2022.000490

[10] Sun, X., Zhu, B., 2020. Development Status and Trends of Self-media Short Videos [in Chinese]. Young Journalist [in Chinese]. 20, 92 – 93.

[11] Zhang, Y., Tang, Y., 2021. Reflection on the Generation Logic and Existing Problems in the Dissemination of TCM Short Videos on the Douyin App [in Chinese]. Contemporary TV [in Chinese]. 3, 90 – 94.

[12] Dong, F., Yu, X., 2024. Factors Influencing the Dissemination Effect of Health Science Short Videos: An Analysis of 50 Verified Doctors' Accounts on the Douyin App [in Chinese]. Communication Innovation Research [in Chinese]. 1, 67 – 87+219 – 220.

[13] Rui, T., 2023. A Simulation Study on the Influencing Factors and Mechanism of Health Science Short Video Dissemination [Master's Thesis, in Chinese]. Huazhong University of Science and Technology: Wuhan, China. DOI: https://doi.org/10.27157/d.cnki.ghzku.2023.004491

[14] Xie, M., 2023. The Effect of Seductive Details on the Dissemination of Health Science Short Videos [Master's Thesis, in Chinese]. South China University of Technology: Guangzhou, China. DOI: https://doi.org/10.27151/d.cnki.ghnlu.2023.003021

[15] He, W., 2021. Research on the Dissemination of Health Regimen Short Videos on the Douyin App [Master's Thesis, in Chinese]. Henan University: Kaifeng, China. DOI: https://doi.org/10.27114/d.cnki.ghnau.2021.001676

[16] Huang, G., 2013. A Study on the Conceptual Metaphors in the English Translation of The Yellow Emperor's Classic of Medicine · Suwen [Master's Thesis, in Chinese]. Fujian Normal University: Fuzhou, China.

[17] Jia, C., 2008. A Study of Conceptual Metaphors in TCM Pathogenesis from the Perspective of Cognitive Science [in Chinese]. China Medicine Guide [in Chinese]. 8, 1141 – 1143.

[18] Jia, C., 2010. Reflections on TCM Theory [in Chinese]. Journal of Beijing University of Chinese Medicine [in Chinese]. 33(7), 441 – 443.

[19] Jia, C., 2011. A Theory Based on Body Experience: An Analysis of Metaphorical Features of TCM Etiology Using "Wind" in "Six Evils" as an Example [in Chinese]. World Science and Technology (Modernization

of Traditional Chinese Medicine) [in Chinese]. 1, 47 – 51.

[20] Jia, C., 2012. A metaphor cognition system of TCM concepts taking water as source domain [in Chinese]. Journal of Beijing University of Traditional Chinese Medicine [in Chinese]. 35(3), 164 – 168.

[21] Jia, C., 2014. Guiding principles for studying the language of chinese medicine based on metaphorical cognition [in Chinese]. Journal of Beijing University of Traditional Chinese Medicine [in Chinese]. 5, 293 – 296.

[22] Li, X., Chen, L., 2017. A Study on Emotional Metaphor Types and Their Cognitive Mechanisms in The Yellow Emperor's Classic of Medicine [in Chinese]. Journal of Xi'an International Studies University [in Chinese]. 25(1), 46 – 50.

[23] Sun, F., 2016. The English Translation of The Yellow Emperor's Classic of Medicine from the Perspective of Conceptual Metaphor [in Chinese]. Shanghai Journal of Translators [in Chinese]. 2, 84 – 88.

[24] Xie, J., Jia, C., 2011. Types and Functions of Metaphorical Discourse in The Yellow Emperor's Classic of Medicine [in Chinese]. Journal of Traditional Chinese Medicine [in Chinese]. 1, 1 – 4.

[25] Yao, C. (Trans.), 2022. Huangdi Neijing (The Yellow Emperor's Classic of Medicine) [in Chinese]. Zhonghua Book Company: Beijing, China.

[26] Lakoff, G., Johnson, M., 1980. Metaphors We Live By. University of Chicago Press: Chicago, USA.

[27] Forceville, C., Urios-Aparisi, E., 2009. Multimodal Metaphor. Mouton de Gruyter: New York, USA.

[28] Kress, G., van Leeuwen, T., 2006. Reading Images: The Grammar of Visual Design. Routledge: London, UK.

[29] Norris, S., 2004. Analyzing Multimodal Interaction: A Methodological Framework. Routledge: London, UK.

[30] Forceville, C., 2009. Non-verbal and Multimodal Metaphor in a Cognitivist Framework: Agendas for Research. In: Forceville, C., Urios-Aparisi, E. (eds.). Multimodal Metaphor. Mouton de Gruyter: New York, USA. pp. 19 – 42.