A Study on the Differences in the Integration of Different Types of Gestures and Language Based on the Modality-Specific Simulation

Hukai Xu
School of Innovation and Development, Central University of Finance and Economics, Beijing, 100081, China
122186540@qq.com

ABSTRACT

The integration of gestures and language is a core topic in the study of multimodal communication. From the theoretical perspective of modality-specific simulation, this paper classifies gestures into four types: indicative, metaphorical, rhythmic, and pictographic. It systematically compares the differences between gesture and language integration in terms of semantic integration depth, temporal process, neural mechanism, and cognitive load, and provides a unified explanation using the modality compatibility hypothesis, the extended dual encoding theory, and the predictive coding framework. Finally, it explores the practical value of this in education and clinical rehabilitation. The research shows that the differences in the integration of different types of gestures are rooted in the different attributes and processing pathways of the modality-specific simulations they activate.

Keywords: Modal Specificity Simulation; Gesture Type; Language Integration; Embodied Cognition; Dual-Stream Model; Modal Compatibility

1. INTRODUCTION

Human communication activities have always been regarded as a comprehensive process of multimodal information processing. Language and gestures are the most representative two-channel expression systems. Traditional language research has always regarded language as the only core channel of communication, and regarded gestures as secondary and meaningless actions. But with the rise of embodied cognition science, more and more researchers have found that there is a deep cognitive coupling relationship between gestures and language [1]. Gestures are not a simple supplement to language, but have complex interactions with language at many levels, such as semantic construction, discourse organization and cognitive simulation. Different types of gestures show great differences when combined with language, and their roots can be traced back to the modality-specific cognitive simulation mechanism of gesture activation [2]. The so-called modality-specific simulation means that when our brain processes different gesture information, it will call the resources of the perceptual motor system corresponding to these gestures, which will form differences in the underlying cognitive mechanism. Understanding these differences can not only help us understand how language and body movements cooperate, but also provide important theoretical support for educational communication, clinical rehabilitation and human-computer interaction [3]. This article wants to systematically talk about how different types of gestures are combined with language, how their cognitive mechanisms are different, and how they can be used in real life, such as education and clinical rehabilitation.

2. THE THEORETICAL FOUNDATION OF MODAL SPECIFICITY SIMULATION

2.1 Embodied Cognition and Modal Specificity Hypothesis

The theory of embodied cognition is the core framework for us to understand how gestures and language are combined. According to this theory, cognition is not an abstract calculation process unrelated to the body, but rooted in the interactive experience between the body and the environment [4]. When our brain processes abstract concepts, it will automatically activate the related sensory motor experience, and this activation is “mode-specific”, which means that the information transmitted by different sensory channels (such as seeing, hearing and touching) is relatively independent in the neural representation in the brain. Modal-specific hypothesis further shows that different modes of information, such as vision, hearing, touch and kinesthetic, have their own favorite processing paths in the brain. When an information pattern is activated, the corresponding sensorimotor cortex will produce corresponding neural activity [5]. This hypothesis gives us an important premise to understand the differences in cognitive processing of different gestures. Because different types of gestures actually use the information of different sensory and motor modules in our brain, when they work with language, the cognitive resources they need to call and the working methods in the brain will definitely be different.

2.2 The Dual Stream Model of Action-Language Integration

In the research on the combination of action and language, dual stream model gives us a well-understood explanation framework [6]. This model distinguishes the different work of ventral pathway and dorsal pathway in the process of combination. Ventral pathway mainly deals with the combination of semantic level, that is, the matching and fusion of action information and semantic expression of language, which requires the use of brain regions such as temporal lobe and inferior parietal lobule. Dorsal pathway, on the other hand, is more involved in the time organization and motion planning of action sequences, which is closely related to the activities of parietal lobe and premotor area. Different types of gestures, due to their differences in semantic content and movement characteristics, do not activate the activation patterns on these two pathways in the same way. Indicative gestures mainly rely on the ventral pathway for spatial semantic integration, while rhythmic gestures more involve the temporal processing of the dorsal pathway [7]. The dual stream model enables researchers to explain from the neural-cognitive level the reasons for the systematic differences in the integration methods and depths of gestures and language.

2.3 Simulation Theory and the Cognitive Basis of Gestures

The simulation theory holds that understanding others' actions and language is not achieved through abstract symbolic operations, but rather through reactivating the perceptual-motor representations corresponding to others' experiences within one's own cognitive system [8]. When an individual observes a gesture, the regions in the brain corresponding to the movement modality involved in the gesture will generate simulation activation. Although this activation is not as intense as the actual execution of the action, it is sufficient to produce a cognitive representation effect similar to the actual action experience. The simulation activation of gestures provides the cognitive basis for their integration with language, as language itself is also understood through the simulation mechanism. When the simulation activation of gestures matches the semantic activation of language in terms of time and content, they can achieve efficient integration. However, due to the differences in the modal attributes of the simulated content activated by different types of gestures, the way and degree of this matching also differ, thereby resulting in the emergence of integration differences.

3. TYPOLOGICAL CLASSIFICATION AND MODAL CHARACTERISTICS OF GESTURES

3.1 Modal Characteristics of Directive Gestures

Directive gestures are one of the most common gesture types in daily communication. Their core function lies in guiding the listener's visual attention through finger or arm pointing movements, pointing to the spatial positions or physical objects mentioned in the discourse. From the perspective of modal specificity, directive gestures mainly activate the visual-spatial representation system. When the speaker extends their fingers and points in a certain direction, the listener's visual-spatial processing system is activated, and the parietal lobe region responsible for spatial encoding generates corresponding neural activity. This activation has a high spatial specificity, and the direction and distance information pointed to will be precisely mapped into the spatial representation of the cognitive system. The semantic content of directive gestures is relatively simple and direct, usually corresponding to the demonstrative pronouns or spatial adjectives in language. Therefore, their integration with language mainly occurs at the spatial semantic level.

3.2 Modal Characteristics of Metaphorical Gestures

Metaphorical gestures refer to those gesture types that express abstract concepts or metaphorical meanings through body movements. Their most obvious feature is that the semantic content of gestures does not directly come from the literal meaning of the action itself, but is mapped from the concrete physical experience to the abstract concept field through the mechanism of conceptual metaphor. For example, move your hands up to express "high emotions", or rotate your hands to represent "the passage of time". From the perspective of modality specificity, the cognitive processing of metaphorical gestures involves the cooperative activation of multiple modalities. Although their external manifestations are information of movement and visual senses, semantic understanding needs to mobilize a wider network of concepts, including emotional experience, time perception, spatial relations and other cognitive fields. Therefore, the modal specificity of metaphorical gestures presents a diffuse cross-modal feature, which makes their integration with language more complicated.

3.3 Modal Characteristics of Rhythmic Gestures

Rhythmic gestures refer to those gestures that match the rhythm and rhythm of language and have repetitive action patterns. These gestures do not have clear lexical meaning or indicative function, but their main functions are to enhance the sense of rhythm of discourse, highlight some information units and adjust the listener's attention state. From the modal characteristics, rhythmic gestures mainly activate the auditory-motor coupling system, that is, the neural network responsible for rhythm processing and action synchronization in the brain. Research shows that the generation and understanding of rhythmic gestures share some neural resources with music rhythm processing, involving structures such as basal ganglia, cerebellum and auxiliary motor area. Because of their low semantic burden, the integration of rhythmic gestures and language mainly occurs at the rhythm and time level, not at the semantic level. This integration mode depends on the synchronization of time structure, not semantic matching.

3.4 Modal Characteristics of Iconic Gestures

Iconic gestures refer to those types of gestures that imitate the shape or action process of an object by changing the position of hands and arms. These gestures are particularly common in deaf sign language systems but also widely present in the daily communication of hearing individuals. The core feature of iconic gestures lies in their high dependence on visual shape information. Speakers use changes in hand configurations to present the contours, sizes, and movement trajectories of objects. From the perspective of modal specificity, iconic gestures mainly activate the visual-shape processing system, involving regions in the occipital lobe and the ventral pathway of the temporal lobe responsible for object recognition. Due to the high overlap of the information carried by iconic gestures in modal attributes with visual language descriptions, their integration with language can occur either at the semantic level or at the perceptual representation level, presenting a unique dual integration mode.

4. ANALYSIS OF DIFFERENCES IN GESTURE-LANGUAGE INTEGRATION OF DIFFERENT TYPES

4.1 Differences in Semantic Integration Depth

The differences in gesture-language integration of different types are first manifested in the depth of semantics. Indicative gestures remain at the level of reference, with gestures pointing corresponding to language indicators in a simple manner, without involving the construction of deep concepts. On the other hand, metaphorical gestures involve conceptual metaphor, and the physical experience activated by gestures needs to be combined with abstract language concepts through metaphorical mapping, which involves the reorganization of conceptual structure. Pictographic gestures are between the two, which mainly rely on visual shape matching, rather than metaphor mapping. Rhythmic gestures do not carry lexical semantics, but only integrate at the level of rhythm and attention regulation. This gradient difference reflects the different positioning of specific modal simulation activated at the cognitive level.

4.2 Differences in Time Progression

In terms of time, indirect gestures usually appear before language, because their functions of guiding attention and establishing common reference need to work before language information processing. Metaphorical gestures are more flexible. They can be synchronized with the language, or they can be a little earlier or later, depending on the complexity of the concept and the processing burden of our brain. The time of Pictographic gestures is related to the complexity of simulated actions. Simple ones can be highly synchronized, while complex ones are prone to time dislocation. Rhythmic gestures are the most stringent, and must be completely synchronized with the rhythm and prosody processing of the language. Any time deviation will destroy the integration effect.

4.3 Differences in Neural Integration Mechanisms

Judging from the way our brain works, the gesture of pointing at things mainly depends on the temporal-parietal junction and frontal afferent gyrus of the left brain, which are responsible for processing space and meaning. Metaphorical gestures also use prefrontal and temporal lobes in the right brain, which shows that two hemispheres need to work together to deal with abstract concepts. On this basis, the occipito-temporal region in the back of the brain will be more active because of the need to process the shape information seen. The rhythmic gestures are mainly the activities of premotor areas, cerebellum and basal ganglia on both sides, and the nerve fluctuation of temporal superior gyrus will be synchronized with the rhythm. These internal differences let us know that different gestures really need the brain to "simulate" the real experience in different ways.

4.4 Differences in Cognitive Processing Load

Indicative gestures have the least burden in dealing with information, and they can even help the language to reduce the burden by repeating expressions. Metaphorical gestures have the biggest burden, because the listener needs to correspond the physical feelings to abstract concepts, which will take up a lot of working memory and brain control resources. Pictographic gestures have a load depending on the complexity of the shape and the matching degree with the language description. Rhythmic gestures, although not involving deep semantic processing, require the continuous occupation of attention resources for maintaining time synchronization.

5. THE THEORETICAL EXPLANATION OF INTEGRATING DIFFERENCES FROM THE PERSPECTIVE OF MODAL SPECIFICITY

5.1 The Modal Compatibility Hypothesis

Modal compatibility hypothesis holds that the integration efficiency is the highest when the modal representation attributes of gestures and languages are compatible. Both indicative gestures and spatial language can activate visual-spatial modes, so they are highly compatible and efficient to integrate. Metaphorical gestures will activate specific sports experience, while language activates abstract semantics. These two modes are incompatible and need additional mapping and transformation. Rhythmic gestures rely on the special compatibility in time structure between movement rhythm and voice rhythm. The efficiency of pictographic gestures depends on the degree of modal overlap between visual shapes and language descriptions. This hypothesis attributes the integration difference to the degree of modal matching.

5.2 The Extension of the Dual Coding Theory

Extended dual coding theory believes that language and gestures form two sets of different expression codes. When their modal attributes are consistent, they will form an integrated code to achieve the best understanding and memory effect; When they are inconsistent, they will form an associated code, which requires additional cognitive operations to establish the connection. Indicative gestures and language will form an integrated code because of the same spatial mode, while metaphorical gestures will form an associated code because of the modal difference between concrete experience and abstract concepts. These differences not only affect immediate understanding, but also affect the maintenance of long-term memory, which provides a theoretical basis for the role of gestures in education.

5.3 The Integration Differences under the Predictive Coding Framework

Predictive coding theory thinks that our brain is like a machine that can predict. Whether it is easy to integrate depends on the “error” of the forecast. Language will give us a prediction in advance, such as indicative gestures, which are very consistent with the prediction of language, so the error is small and the integration will be smooth. But figurative gestures, which often have abstract meanings beyond the expectation of language, will produce great prediction errors, and our brains need to spend more effort to deal with them. The error of rhythmic gestures mainly appears in time. If the rhythm is not right, it will produce obvious error signals. This theoretical framework interprets the difference of integration as the size and type of prediction error, which provides us with a new perspective to understand the dynamic process of integration.

6. CONCLUSION

From the perspective of modal specificity simulation theory, this paper systematically studies the combination of different gestures and languages and their cognitive mechanisms. It is found that there are obvious differences among indicative gestures, metaphorical gestures, rhythmic gestures and pictographic gestures in terms of semantic combination depth, time course, neural mechanism, cognitive load and so on. The root of these differences lies in that different types of gestures activate different modal specificity simulations in attributes, complexity and processing paths. The expansion of modal compatibility, dual encoding theory and predictive coding framework provide complementary theoretical explanations for these differences. With the continuous progress of brain imaging technology and computational modeling methods, it is expected that the dynamic process of integration of different types of gestures and languages will be revealed on a finer neural spatio-temporal scale in the future, thus providing a more solid foundation for the theoretical development and application practice of multimodal cognitive science.

REFERENCES

[1] Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. Psychological bulletin, 137(2), 297.
[2] Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron, 41(2), 301-307.
[3] Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. Cerebral Cortex, 17(10), 2322-2333.
[4] Barsalou, L. W. (2008). Grounded cognition. Annu. Rev. Psychol., 59(1), 617-645.
[5] Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. Journal of physiology-Paris, 102(1-3), 59-70.
[6] Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature reviews neuroscience, 8(5), 393-402.
[7] Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. Spatial cognition and computation, 5(4), 307-331.
[8] Gallese, V., & Lakoff, G. (2005). The brain’s concepts: the role of the sensorimotor system in reason and language. Cognitive Neuropsychology, 22(3), 455-479.