Research on Robot Interaction Design Based on Embodied and Disembodied Interaction Grammars

Jingyan Qin[(✉)], Xinghui Lu, Yanlin An, Zeyu Hao, Daisong Guan, Moli Zhou, and Shiyan Li

School of Mechanical Engineering,
University of Science and Technology Beijing, Beijing, China
qinjingyanking@foxmail.com

Abstract. With the increasing complexity and variety of robotic agents in VR/AR/MR IoT and artificial intelligence and other intelligent environments, traditional interaction grammar and interaction design principles are facing more and more challenges. Based on the theory of embodied interaction and disembodied interaction in cognitive science, this paper combines field research and user in-depth interviews, usability testing, SWOT analysis, SPSS questionnaire analysis, high-fidelity prototype construction, usability testing and other research data to study the interaction grammar of embodied interaction and disembodied interaction. By designing the prototype of the robot, three typical Chinese families are selected to conduct prototype tests of embodied interaction and disembodied interaction. Then on this basis, affordance, Interactive Behavior, Interactive Feedforward of embodied interaction and Mapping Relations of Symbols, Semantics and Interactive Feedback of disembodied interaction and understanding of the construction of interactive grammar are established. This study has a certain effect on embodied interaction and disembodied interaction between human and computer interaction.

Keywords: Interactive grammar · Embodied cognition ·
Disembodied cognition · Robot interaction design

1 Introduction

Based on the disciplines of Phenomenology of Perception and Cognitive Science, George Lakoff, Mark Johnson, Dourish, A Clark, Edmund Gustav Albrecht Husserl, Martin Heidegger, Maurice Merleau-Ponty [1–6] and other scholars from cognition science and computer science conducted research on embodied cognition, disembodied cognition and embodied interaction. It can be seen from the above literature that the body exists in space, and participates in the visible expression form of the human subject's intention to the object as a medium for perceiving the world. The subject exchanges and interacts with the object through the sense of existence, and communicates with the traffic sympathy and so on, and constructs the concept of embodied interaction. Through tele-presence of information communication, disembodied cognition enables human consciousness to be separated from cognitive vectors, and even

creates a sense of coexistence co-presence between real and virtual bodies through virtual reality and mixed reality. At the same time, AI can also concentrate group wisdom in a carrier, which produces the separation of consciousness from subject. And disembodied cognition produces disembodied interaction in the process of observing perspective of exchange consciousness, information dimension of exchange consciousness and energy of sympathetic consciousness between subjects and objects. If the interactive representation layer, the operation layer and the container layer are separated from each other, the disembodied interaction will take a leading role, and if the representation layer and the operation layer tend to be one in the entity interaction interface or the augmented reality environment, thus the embodied cognition will become dominant. Aiming at the human-computer interaction environment of robotic agents, this paper proposes the practice of interactive design using the interactive grammar of exposed interaction and disembodied interaction. By designing the prototype of the robot, three typical Chinese families were selected to perform the prototype usability test of the embodied interaction and disembodied interaction, so as to establish the mapping relationship of the embodied interaction and disembodied interaction and the understanding of the interaction grammar.

In this study, we try to use the interactive grammar constructed by embodied cognition and disembodied cognition to guide the design process of the agent and human-robot interactive user experience. In the Chinese family form, human-computer interaction content is designed for various needs of users in different scenarios, and prototype testing is carried out in a real environment. We have designed two different forms of agents. By recording the cognitive perception and interaction behavior of the user under the embodied cognition, the interactive grammar design under the embodied cognition and disembodied cognition is very important for the design of human and agent.

Considering the difference in the way in which the embodied cognition and disembodied cognition record and analyze data, we observe the user's susceptibility and functional requirements for the agent under the disembodied interaction experience of the existing agent. Through usability testing and in-depth interviews, we explore the embodied interaction and disembodied interaction of users and agents. Through the iterative tests, it proves that the function requirements and emotion experience of the agent match the interactive grammar under the embodied and disembodied cognition proposed in this study.

2   Experiment Design for Human-Robot Embodied Interaction and Disembodied Interaction

2.1   Determination of Test Content

In this study, 160 valid questionnaires were collected for the intelligent product demand survey. The questionnaires covered users' needs, product pain points and opportunity point troubleshooting, interactive form preferences, and users' trustworthy AI. The age covers 18–71 years old, 20 provinces and cities in China, 4 consumption levels, and the occupation of 19 fields, the proportion of men and women is balanced.

The survey found that users have needs in ten aspects: emotional companionship, entertainment, family interaction, social interaction, child growth and education, health management, life management, and safety monitoring. Based on the above requirements, we selected 28 scenes tasks cards, including the functions, content, and interaction of the agent.

The survey analyzes the core life claims of current families are reflected in ten aspects:

1. Accompany to get rid of loneness and ease monotony
   Accompanied by people and things, ease monotony, boring, and bring emotional comfort. Especially for solitary people and the elderly.
2. Entertainment
   Take a variety of ways to enjoy leisure and seek physical and psychological satisfaction. Especially living alone, a family of two, a family with children.
3. Family interaction
   The interaction between people and the relationship between family members. Especially between husband and wife, family and parent-child interaction.
4. Social interaction
   The interaction between people and people maintains social relationships with others. Especially living alone, the elderly.
5. Child growth and education
   Healthy growth and good education of children are one of the main demands of children's family life. Especially family and parent-child interaction.
6. Health Management
   Health is a strong concern for all families, including physical & mental health. Especially the elderly.
7. Life management
   Managing all kinds of trivial matters in life is the basic need in family life.
8. Safety monitoring
   Family security, keep home safe when not at home or sleeping.
9. Energy saving and environmental protection
   People's awareness of environmental protection has gradually increased, and all aspects of life have begun to advocate conservation to create a healthy and green living environment. (saving electricity, water, etc.)
10. Price threshold
    Due to the large and full price of the agent, for some families, the price is relatively high, and the price and function should be considered. Especially for solitary people.

The classification of needs based on four-quadrant maps of appeals is performed (Fig. 1), urgent and important ones include family interaction, companionship and ease of monotony, life management. Urgent but not important ones include child growth and education. Important but not urgent ones include health management, safety monitoring, energy saving and environmental protection. Those neither urgent nor important include entertainment, social interaction.
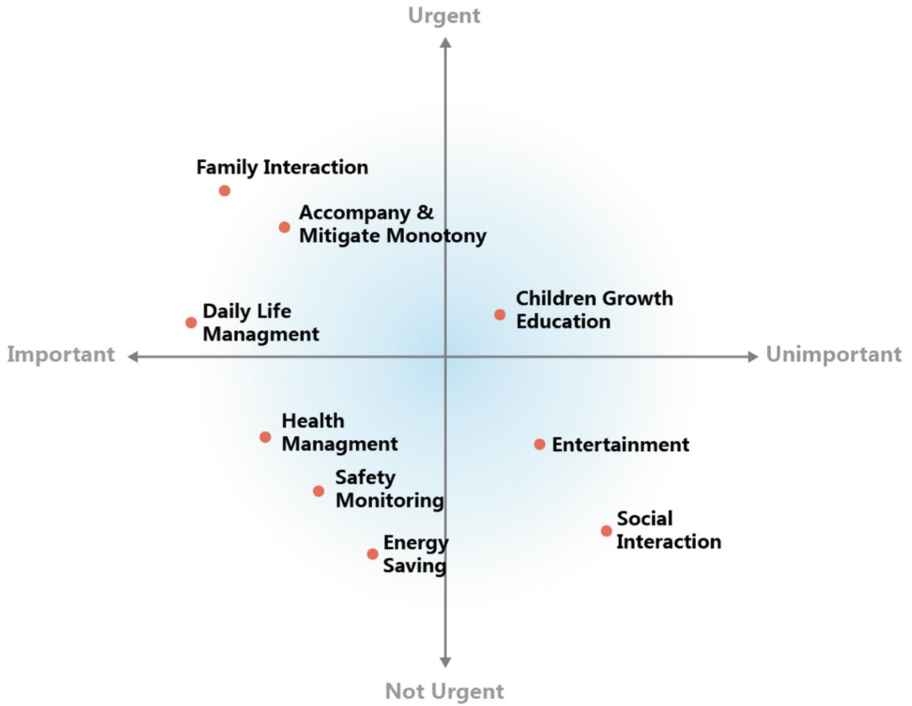
Fig. 1. Touchpoint four quadrant analysis

At present, the core pain point of home users lies in that functional operation feedforward and interactive disembodied cognition required by embodied cognition cannot be seamlessly transformed into multi-modal communication protocols by content-based speech feedback, and it is impossible to establish a mapping transformation relationship of manipulated interaction and disembodied interaction through interactive grammar.

In view of 9 field surveys, 22 in-depth user interviews, SWOT analysis results of 50 existing robot competitors and agent-related products, 9 kinds of robots and agents (including pudding Pudding Beanq, FABO, DuSmart Speaker, Xiaoduzaijia, Xiaodu robot, TmallGenie, Xiaoai Smart Speaker, Atals, Luka) (Fig. 2) are selected to perform usability test and build prototype construction and design experiment for a robotic agent that conforms to embodied interaction and disembodied interaction. Through the anthropomorphization of the agent, the roles of the six robots are designed, and 16 typical robot agent usage scenarios are designed and planned to perform the user experience test of embodied interaction and disembodied interaction.

Fig. 2. Robot agent competitive product analysis

- **The feedforward information input and analysis of the function of the agent requires the agent to satisfy the following conditions**

1. Users want the agent to know enough about themselves
2. When considering the functional design point of the agent, the active interaction that relies too much on the agent will cause user inertia.
3. People have the ability to feel, control and choose, family robots can't make decisions instead of humans
4. Does the agent need to have mobile function? Which includes:
   (1) The function of providing services through mobile
   (2) Follow - emotional needs

- **Function aspects of the agent**

1. Individual
   (1) From the perspective of family space: living room, bedroom, kitchen, bathroom (information receiving object: object):
       The family has multiple spaces (functions in different spaces), different spaces, different sizes, and different task points.
   (2) From the perspective of human existence (information recipient: person):
       There are multiple scenes in the family, and scene (event) recognition is a problem.
   (3) From a personal perspective: companionship, entertainment, education – emotion:
       Different users have different function and emotion needs and launch targeted services.
2. Ecosystem
   (1) The relationship between the agent and other smart devices (smart home + mobile phone):
       The existing agent cannot handle the relationship with the mobile phone, the information did not open, and the closed loop could not be formed. The intelligent body's control of the smart home has not achieved intelligent full automation and needs to be improved.

(2) Relationship of family members:
    The biggest problem with family members in communication is that the family time information is asymmetrical, and they do not understand what the other party is doing, which may lead to emotional injuries and misunderstandings. Family members have less time to communicate.

- **Content aspect of the agent**

1. The information content presented by the agent are too scattered, and the user cannot extend the data into knowledge graph, which is not convenient for information inquiry.
2. The relationship between the mobile phone and the home agent, whether the home agent can be replaced by the mobile phone, what kind of information is biased to be presented by the mobile phone, and what information tends to be presented by the smart body.

- **Performance and discovery points in terms of interaction forms include**

1. Family members have different roles and different tonality, which will result in different feedback mechanisms and content feed-forward mechanisms.
2. Users are more inclined to language input, and the choice is more inclined to touch screen selection.
3. The existing agent interaction mode is single, and the user prefers to have a physical interaction with the agent.
4. Users have more new expectations for the content output form of the agent.
5. Embodiment: Different users have different requirements for the form of the agent (appearance: color, size, material; system presentation: expression, person setting, sound, etc.) (mental model)

- **Roles of the agent**

Based on the above research and analysis, we define the family agent as the assistant, the housekeeper, the friend, the bodyguard, and the nanny at the functional level. This includes the reminder function of the assistant, the auxiliary solution to other problems in life, the management function of the housekeeping type, the function of sharing the time of the friend's entertainment, the home security of the bodyguard, the guardianship, care and help function, encyclopedia-style advisory guidance like teachers. For the interactive function, the prototype is designed for the embodied interaction and designed the disembodied interaction for the interactive content, and the prototype is designed and tested in 16 family scenarios.

The embodied interaction includes: the affordance display of functions brought about by unconscious behavior, the interaction between humans and robots, including

the physical operation of touch screen control, beckoning, hugs, touching, facial expressions, etc., and interactive feedforward information for embodied interaction. The disembodied interaction specifically includes: an interaction symbol for the intelligent entity environment and digital information in the interactive interface, an interaction semantics constructed according to interactive grammar rules such as metaphor, implicit metaphor, metonymy, and the feedback function of the activated disembodied interaction based on the feedforward information of the exposed interaction.

2.2   Scene Test

For the physical test environment, we chose to test in the real home scenario (Fig. 3), which shows the relative relationship between users, agents and environment in real scenes. In this scenario, the user can be motivated by the original embodied cognition. In the prototype design of the agent, we designed two different sizes of agents to explore Cognitive Efficiency of Users on the Ability of Agents by recording the interaction behavior and distance between agent and user, the expression of agent and user's emotion, and perform mapping between the semantics of disembodied interaction and the behavior of exposed interaction in interactive grammar. Prototype testing of three typical Chinese families [one core family (second generation), one direct family (three generations), one single-person young family] is performed using the scenario of the agent, three typical family members including four office workers, 1 housewife, 1 primary school student, 1 university student, 1 old man retired at home.

Fig. 3. People and agents in a test environment

3   Result Analysis

3.1   Analysis of the Results of the Embodied Interaction of the Agent

- **Agent appearance**

The appearance of the agent will bring about a difference in function. A large agent has a sense of presence while a small agent has a sense of family integration. The user's interaction behavior is different under the agents affordance cues. When users interact with large-scale agents, they are more of the click behavior of visual interface and the language behavior of exposed interaction under disembodied interaction. When users interact with small-sized agents, they mainly focus on exposed interaction behaviors, such as touching and tapping, hug, etc.

- **Embodied interaction behavior between human and agent**

When people interact with each other, various implicit mechanisms are used to share information about their state and interaction state. The mechanism by which the physical state of the body is changed is called a body language. Body language provides clues about personal emotions, emotions, and mental states, which can perform information communication by posture, gesture, facial expression, eye behavior, contact with body-related behavior and pronunciation behavior [7]. Studies have shown that people interact with robots in social network models [8]. When faced with complex inanimate objects, people often use social models to understand and predict their behavior [9]. As a result, such model is applied to embodied interaction between human and robot agents. In the process of interaction between humans and agents, body language behavior plays an important role in human-robot interaction. The human body language signal sends additional clues that can be observed, which characterize the user's mental and physiological state and his intentions, thus making the human-robot interaction intuitive and efficient. The signals and gestures of human body language appear in the form of clusters. The various channels of body language are semantically ambiguous and need to be understood in specific contexts. The most frequent interaction between humans and agents is the contact behavior between the two, including human touch, limb collision, greeting, hug and guidance (Fig. 4), the semantics of these interactions depend on the context, including the nature of the relationship and the way it is performed. It requires the agent to perform accurate user intention analysis and convey corresponding natural output feedback.

3.2   Analysis of the Results of the Disembodied Interaction of the Agent

- **Agent Symbols**

When designing the disembodied interaction of robotic agents, we can learn from human body language behavior and use this natural behavior to express unique personality and characteristics, which will help the strong tie between humans and robots, such as friendship and trust. By creating a symbolic system of facial expressions, body language, functional operations, and content semantics, the agent constructs human-
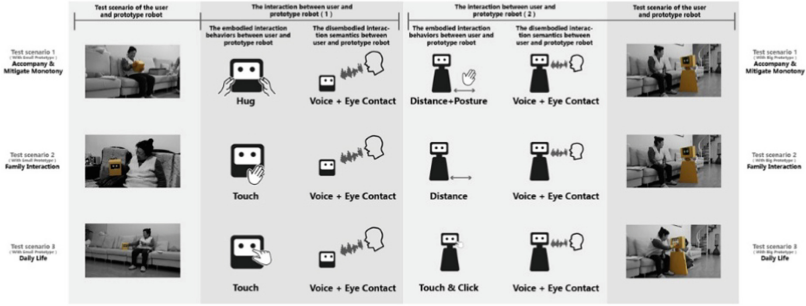
Fig. 4. The interaction between users and two different robot prototype

computer interaction in the interface performance, and correspondence with the level of interaction behavior. By symbol, signifier and sign, the language content and grammar standards of human-computer interaction are established, as show in Fig. 5.
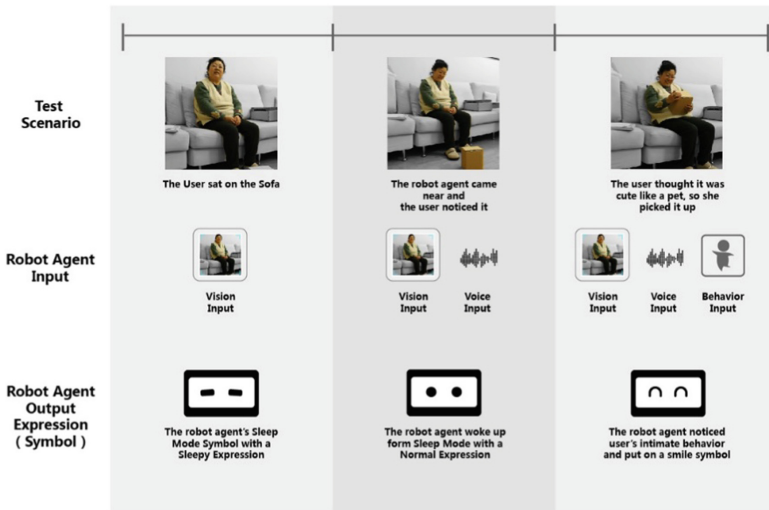

Fig. 5. The symbols of disembodied interaction between users and robot

1. Expressions

   Facial expressions are the main source of emotion. At present, most of the research focuses on six main emotions, namely, anger, sadness, surprise, happiness, fear and disgust. The expression design of the agent can be used as a reference for conveying emotions. It can be matched in view of input for user interaction, so as to express human personality characteristics to enhance the natural and real sense of human-computer interaction.

2. Postures

The posture is used to convey information about attitudes. By designing the interactive feedback of the agent, the degree of attention paid by the user and the degree to which the agent responds to the person can be determined. The survey shows that the user prefers the agent to have a forward-looking behavior when performing interactive feedback. At the same time, the posture can also reflect the intensity of emotional state, for example, the weakness of the body is connected to the sad mood, and the nervous posture is connected to the angry emotion. Therefore, when designing the expression of the agent, it can also be combined with the body posture to increase the natural feeling of human-computer interaction.

3. Eye-contact behavior

Eye-contact communication plays an important role in interpersonal communication. The exchange of information between people always starts with eye communication. Eye behavior plays an important role in the transmission of information. Therefore, the eye behavior design of the agent is more conducive to emotional expression. Through eye behavior, an agent can express its "emotional" state, passing on more information that languages and postures cannot accurately represent. Eye behavior is also an indicator of interest, attention or implicature. In the process of uninterrupted eye contact between the agent and the person, the user can better understand the intention of the agent and thus exchange information.

- **Semantics of agent**

Through interactive information content, the agent establishes the communication of human-computer interaction in meaning, including functional operation and language meaning. In the design of man-machine dialogue mechanism, because the mental model of human intelligence is different from the cognitive computing of artificial intelligence, and the content meaning of FBS (function behavior structure) and CMR (content model relationship) is different, then words and semantics are brought into important attention. The living data generated by the continuous source of dialogue provides an important basis for human-computer interaction data services. The design of Human-Robot Dialogue follows a semi-open dialogue structure, which can not only satisfy the openness of human dialogue and communication, but also change the functional behavior structure model that the artificial intelligence has no clear task pointing. An example analysis of the two test scenarios of the agent is performed as follows.

1. Scene 6——Child Education and Counseling

Little brother: Hello, I don't know much about this history problem.
Agent: I am here to help you, what is the problem?
Brother: About XXXXX
Agent: OK, I know, I am inquiring for you now.
(after finding the answer)
Agent: OK, I have found the right answer for you, come check it out!

The intelligent body-speaking design for children needs to turn children's education into heuristic education, which uses guided dialogue structure. According to the content

that the user is interested in, the result of the knowledge search of the content is displayed, instead of directly giving the answer, the answer of the question is given step by step to form an interactive dialogue structure, and thus guiding the subject to actively think. The possibility of multiple answers to a question is explored using ways of encouragement and suggestion, thus allowing users to form a model of independent thinking and a unique way to explore their own answers.

2. Scene15——Reminder of sleep
   Agent: It is already 0 o'clock in the morning, you should rest.
   Youth: My work is still not finished, and I will sleep in half an hour.
   [After half an hour]
   Agent: It's been 30 min, is your work done?
   Youth: Still a little bit, but I am too sleepy, help me set a clock at 6 o'clock tomorrow, I will continue to do it tomorrow morning.
   Agent: OK, I have set an alarm clock for you at 6 o'clock tomorrow morning. Go to sleep, good night.

Verbal design of time reminder shall not apply rude tones, or sermon style, but the equal dialogue that users can accept. For example, the user should be informed of the amount of work tomorrow or the benefits of early sleep and warm heart greetings. The language content should not be too much, which may be easy to cause users to resent. In the time of sleep, it can also play related music or use low-frequency white noise to get the user into a sleepy state, and naturally remind users.
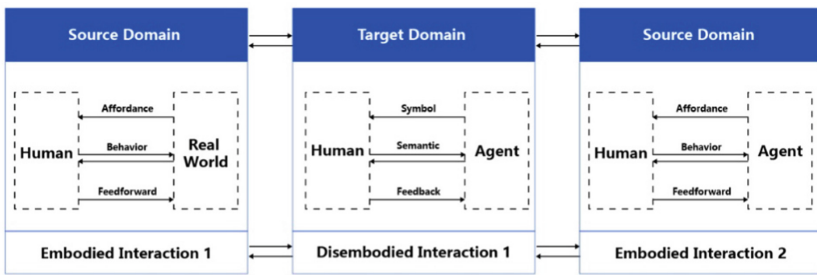
- **Interactive feedback**

The agent provides information feed-forward through the words content of the agent, and the information feed-forward provides a basis for the interactive feedback of the function symbols. Therefore, the agent should provide both the functional design of the weapon type and the content design of the container type. Integrating factors such as functions, interaction behavior, data structure, user vision and requirements, information container content, mental model and interaction model, relationship between human-machine environment, relationship between scene and story, etc., as well as the establishment of interactive grammar, a series of design meta-language and meta-translation systems of the interaction design, product design, service design and ecological design of the intelligent robot agent are constructed.

4   Mapping Relation Between Embodied Interaction and Disembodied Interaction

The disembodied interaction of the agent is extracted by the subject (human) from the object (the real environment) and the object (the real environment) to the subject (human) perception, cognition and behavior (embodied interaction). Advocates of the Embodied interaction view believe that artificial intelligence cannot completely replicate different levels of intelligence, and the biological nervous system that realizes human cognitive ability cannot be completely equivalent to computer hardware systems. Therefore, artificial intelligence does not understand the mindset and sensibility

determined by the body like human intelligence. Through the knowledge graph, the cognitive experience of the human objective body (the body at the level of the biological nervous system) and the body of the phenomenon (the body experienced in the social culture) are associated with the functional requirements of the input. The embodied interaction is important for guiding the agent to do what, how, and why it interacts with people. Using the perception and cognition of the real world under the imposed interaction, the behavior feedback design of the agent is constructed to constitute the affordance demonstrativeness of the agent. In the process of human-robot interaction, the behavior of the agent stimulates the experience of the past and the new embodied interaction with the agent. The mapping relationship of the embodied and disembodied interaction is shown in Fig. 6.

Fig. 6. The mapping relationship between embodied interaction and disembodied interaction

Since between affordance, behavior and symbols in the embodied interaction with symbols, semantics, and feedback in the disembodied interaction there is lack of mapping relationship constructed by the interaction grammar, the human-robot interaction causes insufficient interaction performance and user experience. The open learning of artificial intelligence provides feed forward information for the intelligent agents. The detection framework of the intelligent environment is more efficient. The compact and efficient depth CNN feature helps the agent to perform the identification and judgment of the outgoing interaction and the spontaneous weak supervision or unsupervised learning of the symbolic semantics, so as to form the knowledge graph of interactive semantics. Quantify-self forms the robust target representation control of the interactive object, and provides the Basis of Three-Dimensional Target Detection of Situation Reasoning in Social Network Computing of disembodied interaction, target instance segmentation of interactive disembodied interaction and embodied interaction of interaction objects to demonstration of embodied interaction, making the feedforward of the imposed interaction and the feedback of the disembodied interaction causally related to each other.

5 Conclusion

Through the questionnaire survey, in-depth interview and agent prototype usability test, this study draws the interactive content and functions mutually carried by embodied interaction and disembodied interaction. Affordance, behavior, feed-forward in embodied interaction and symbol, semantic, and feedback in disembodied interaction form a complete interactive grammar system. Then it performs Subject-Object Embodied Representation and Subject Cognition through body language, facial expressions, sentence intonation, interactive entities, and constructs systematic mapping of the origin domain between embodied cognition and disembodied cognition. The interaction grammar (interaction behavior pattern recognition; interaction model and information architecture, function and content relationship, etc.) is used to design the interaction relationship between human and robot agent to further verify the mapping relationship among the three elements in the event of embodied interaction and disembodied interaction.

References

1. Lakoff, G., Johnson, M.: Philosophy: The Embodied Mind and Its Challenge to Western Thought. Beijing World Publishing Corpotion, Beijing (2018)
2. Dourish, P.: Where the Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2001)
3. Clark, A.: An embodied cognitive science? Trends Cogn. Sci. **3**(9), 345 (1999)
4. Edmund, G.A.H.: Phenomenology of Intersubjectivity, p. 57. Maldinus Neihov Publishing House, Hague (1973)
5. Dreyfus, H.L.: Being-in-the-world a commentary on Heidegger's being and time, division I. Philos. Lit. **15**(2), 373–377 (1991)
6. Merleau-Ponty, M.: La Phénoménologie de la Perception. Gallimard, Paris (1945)
7. Xu, G., Tao, L., Di, H.: Body Language Understanding for Human Computer Interaction, pp. 16–18. Publishing House of Electronics Industry (2014)
8. Breazeal, C.: Emotion and sociable humanoid robots. Int. J. Hum. Comput. Stud. **59**(1–2), 119–155 (2003)
9. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York (1996)