Addressing the Symbol Grounding Problem in VR

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

Makoto Toyota
T-Method
Chiba, Japan
Email: pubmtoyota@mac.com

Katsuko T. Nakahira
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

***Abstract*— A Virtual Reality (VR) environment presents objects that the user perceives and interacts with. It then transitions to the next state, reflecting the content of the interaction that has occurred as a result of the user's perception of the objects. For the interaction between the VR environment and the user to continue seamlessly, the meanings assigned to the objects in the VR environment by the creator of VR applications and the meanings held by the user experiencing them must be consistent. In this study, we propose a method to realize seamless interaction between VR environments and users by considering the objects presented by VR environments as symbols and capturing the relationship between the meanings they contain and the meanings held by the users who experience them through the symbol grounding problem, which is regarded as a challenging issue in the field of artificial intelligence. Based on the Mode Human Processor with Realtime Constraints (MHP/RT), which is a cognitive architecture that can deal with action selections in everyday environments, we focus on the fact that the content of human action selections is based on memes that are handed down from generation to generation and should provide a basis for his/her understanding of the situation of the surrounding world, and suggest that the symbol grounding problem can be solved by observing and identifying memes.**

Keywords*— *The Symbol Grounding Problem; Meme; Virtual Reality; Artificial Intelligence; MHP/RT; Structured Meme Theory.

I. INTRODUCTION

Humans acquire information about the outside world through the five senses, and select and execute appropriate actions for the situation at hand by operating the Perceptual, Cognitive, and Motor (PCM) processes. Perceived information is encoded and represented as symbols in the perceptual process, which makes it possible to think in the cognitive process. In the thinking process, memory is used to successively transform the symbols into new representations. Part of the result of thinking gives a representation of a sequence of actions that can be performed in the motor process. Based on the idea that human intelligence can be captured by thought processes that manipulate symbols, Newell proposed the Physical Symbol System as a theory of human intelligence [1]. This idea provided the foundation for Soar [2][3], which is one of the successful cognitive architectures.

The information that is input to the perceptual process through the sensory organs has its source in the real world. According to the sequence of actions represented by symbols generated through the PCM processes, actions are performed in the real world and the real world is updated. When we try to artificially realize such interactions that humans perform in the real world using a physical symbol system, the system must have the ability to link the symbolic representations to the references in the real world and to acquire meaningful understanding from interactions with the environment.

The realization of this capability is a fundamental challenge in Artificial Intelligence (AI) research and is referred to as the Symbol Grounding Problem (SGP) [4]. It concerns the ability of a machine to connect its symbolic representations to real-world references and acquire meaningful understanding from its interaction with its environment. In other words, it is about how machines can understand and represent the meaning of objects, concepts, and events in the world. Without the ability to ground symbolic representations in the real world, machines cannot acquire the rich and complex meanings needed for intelligent behavior such as language processing, image recognition, and decision making. Addressing the SGP is crucial to building machines that can perceive, reason, and act like humans.

One of the environments in which humans interact is a Virtual Reality (VR) environment. In a VR environment, users can interact with artificial 3D visual environments or environments involving other sensory modalities using computer modeling and simulation. VR applications immerse the users in a computer-generated environment that simulates reality. In a VR environment, user interaction proceeds through user-perceivable objects provided by VR applications. The meaning that the user gives to the objects generated by the VR applications determines how the user interacts with the objects. The VR applications can achieve a seamless interaction by having the ability to appropriately handle the meanings given by the user to the objects. Here, it can be seen that the SGP is not unrelated to the realization of seamless VR environments. In this paper, we propose a method to deal with the SGP in VR.

This paper is organized as follows. Section II presents the framework for dealing with interaction via objects. Section III describes the interaction between self and an object, how to capture the SGP in AI from the interaction perspective, and how to capture the interaction between self and VR. Section IV suggests how to generate VR environments that could guarantee symbol grounding.

II. INTERACTION VIA OBJECTS

In this section, the aspects of individual human interaction through objects are classified in Section II-A and the perceptual, cognitive, and motor processes and memories of each human who interacts with the object are described in Section II-B.

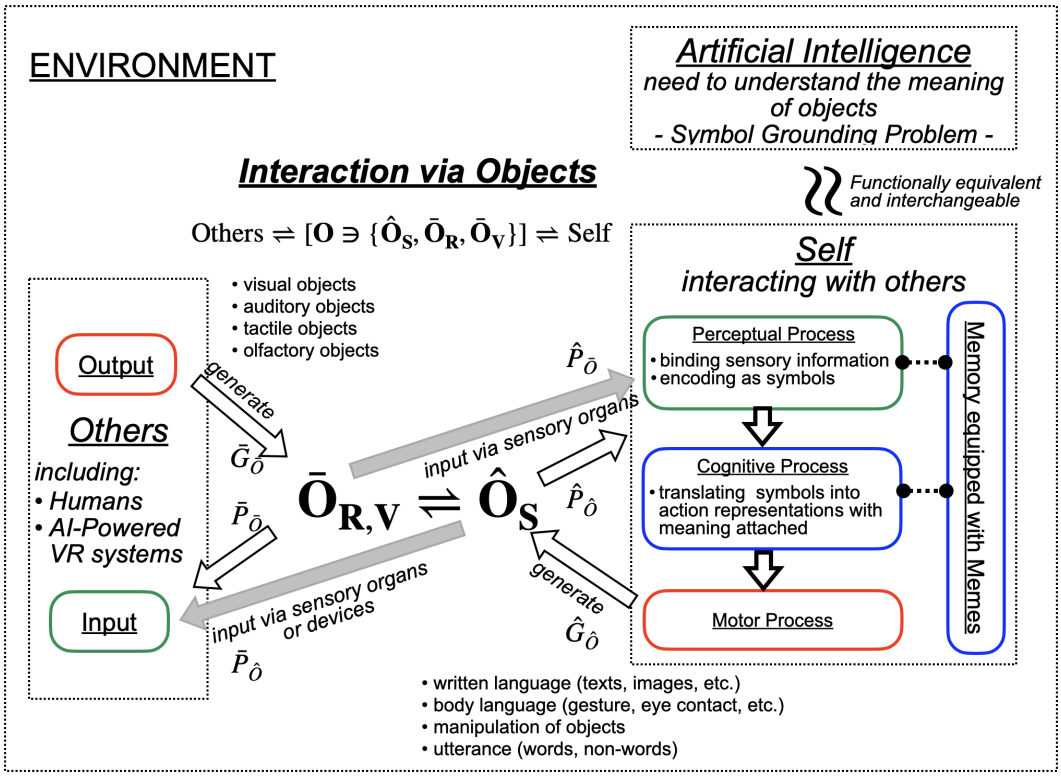

Figure 1. Interaction via objects

A. *Interaction Types*

Individuals live their daily lives interacting with a large number of objects that exist around them. Objects are classified according to whether they are directly or indirectly interacted with by self and by whom they are created. There are three types of objects defined as follows:

- $\hat{\mathbf{O}}_{\mathbf{S}}$: Real objects that are directly generated by self, e.g., utterance, written text, drawings, gestures, modeling, etc.
- $\bar{\mathbf{O}}_{\mathbf{R}}$: Real objects generated by other human beings with whom self is directly interacting.
- $\bar{\mathbf{O}}_{\mathbf{V}}$: Objects generated by a machine or other human beings with whom self is not directly interacting.

These three types of objects are collectively represented as $\mathbf{O}$ ($\ni \{\hat{\mathbf{O}}_{\mathbf{S}}, \bar{\mathbf{O}}_{\mathbf{R}}, \bar{\mathbf{O}}_{\mathbf{V}}\}$). In Figure 1, $\mathbf{O}$ is placed in the center, "Self" interacting with $\mathbf{O}$ on its right side, and "Others" interacting with $\mathbf{O}$ on its left side. On top of "Self" is an "Artificial Intelligence" that is functionally equivalent to self and can be replaced. Self operates the PCM processes to generate the objects $\hat{\mathbf{O}}_{\mathbf{S}}$ ($\hat{G}_{\hat{O}}$). Self also perceives them ($\hat{P}_{\hat{O}}$). The objects self perceives include the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ generated by others ($\hat{P}_{\bar{O}}$). Meanwhile, others generate the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ ($\bar{G}_{\bar{O}}$). If the other is a human being, it runs the PCM process, which is equivalent to the one self runs, to generate the objects $\bar{\mathbf{O}}_{\mathbf{R}}$ ($\bar{G}_{\bar{O}_R}$). Machines run their own generative mechanisms to produce the objects $\bar{\mathbf{O}}_{\mathbf{V}}$ as output ($\bar{G}_{\bar{O}_V}$). The input to the others are the objects $\bar{\mathbf{O}}_{\mathbf{R},\mathbf{V}}$ generated by the others themselves ($\bar{P}_{\bar{O}}$), or the objects $\hat{\mathbf{O}}_{\mathbf{S}}$ generated by self ($\bar{P}_{\hat{O}}$). In summary, interaction via objects can be expressed as follows:

$$\text{Others} \rightleftharpoons [\mathbf{O} \ni \{\hat{\mathbf{O}}_{\mathbf{S}}, \bar{\mathbf{O}}_{\mathbf{R}}, \bar{\mathbf{O}}_{\mathbf{V}}\}] \rightleftharpoons \text{Self}$$

B. *Interaction between Self and* $\mathbf{O}$

The interaction between self and $\mathbf{O}$ is performed by the PCM processes that the self runs, and by the memory processes that are used by the PCM processes and updated as a result of the execution of the PCM processes. This section provides an overview of the PCM and memory processes based on the Model Human Processor with Real-Time Constraints (MHP/RT), a cognitive architecture that can simulate everyday action selections [5][6][7].

1) The PCM Process: When interacting with objects in the environment, humans respond to physical and chemical stimuli emitted from the objects by sensory nerves located at the interface with the environment and take in environmental information in the body. Figure 2, adapted from [8, Figure 1] with modification, shows the PCM process, based on the MHP/RT cognitive architecture [6][7], by which environmental information is taken into the body via sensory nerves, processed in the brain, and then acted upon by the external world via motor nerves. This process uses memory, which is modeled as the Multi-Dimensional Memory Frame. It consists of the Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-Dimensional Memory Frame. The Perceptual-Multi-Dimensional Memory Frame overlaps with the Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame. This

 ISBN: 978-1-68558-266-1

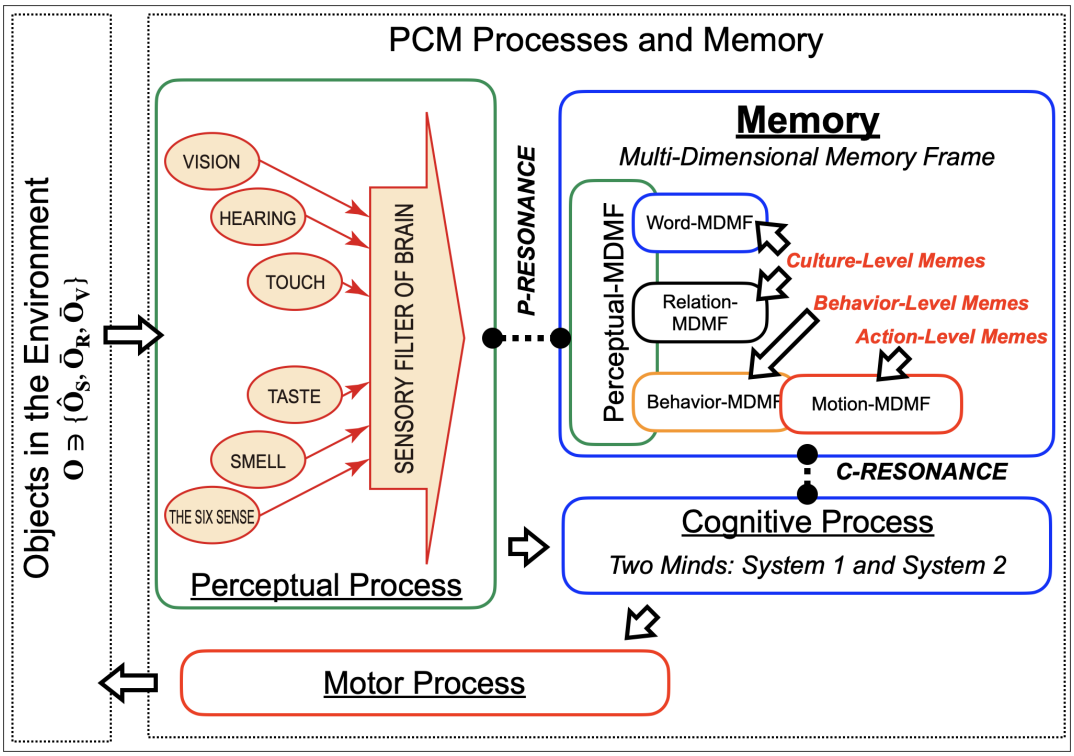

Figure 2. PCM process and memory

is the unique and indispensable configuration of memories defined by the Multi-Dimensional Memory Frame for spreading activation from the Perceptual- to Motor-Multi-Dimensional Memory Frame, which connects perception with bodily movements.

Perceptual information taken in from the environment through sensory organs *resonates* with information in the Multi-Dimensional Memory Frame, which is called P-Resonance. In Figure 2, this process is indicated by the symbol •—•. Resonance occurs first in the Perceptual-Multi-Dimensional Memory Frame and activates the memory network. After that, the activation spreads to the memory networks that overlap the Perceptual-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame. In cognitive processing by Two Minds [9][10], conscious processing by System 2, which utilizes the Word- and Relation-Multi-Dimensional Memory Frame via C-Resonance, and unconscious processing by System 1, which utilizes the Behavior- and Motor-Multi-Dimensional Memory Frame via C-Resonance, proceed in an interrelated manner. The motor sequences are expressed according to the Motor-Multi-Dimensional Memory Frame. The memories involved in the production of actions are updated to reflect the traces of their use process and influence the future action selection process.

2) Memory and Memes: When the PCM process is running, the contents of Perceptual-Multi-Dimensional Memory Frame are updated in response to the perceptual process, those of Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame are updated in response to the cognitive process, and those of Motor-Multi-Dimensional Memory Frame are updated in response to the motor process. Alternatively, the memory system can be viewed from the perspective of memory use. The integrated sensory information first activates the Perceptual-Multi-Dimensional Memory Frame; then the activation spreads to the Word-, Relation-, and Behavior-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame bound to the motor nerves. The basis of behavior is *imitation*; do as what you see. Therefore, behaviors that can be imitated across generations are preserved as sustainable behaviors. In this way, we can organize the Multi-Dimensional Memory Frame, which is used by the PCM processes and updated by their execution, in terms of *memes* that can be inherited across generations [11].

Word is considered the archetype of meme [12]. Words, i.e., symbols, are gradually incorporated into the environment in the form of *thesauruses*, i.e., lists of words in groups of synonyms and related concepts; followed by incorporation of languages used for person-to-person communication, *individual languages*, which might include not only direct but also metaphorical uses; and lastly incorporated are languages used in cultural contexts, *cultural languages*, in which appropriate understanding of common sense that has been established in the specific community, is essential for successful communication. These three forms circulate among people and persist from generation to generation [13].

Thesauruses, individual languages, and cultural languages increase their complexity in this order in terms of the patterns they are linked with the objects in the environment. The-

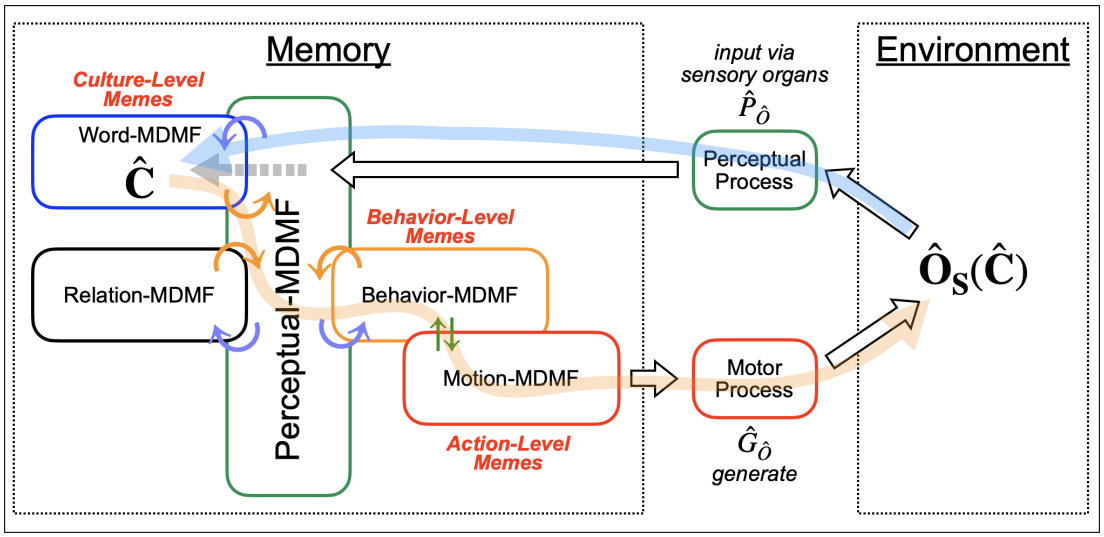

Figure 3. Symbol grounding when the self generates an object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ that embodies the concept $\hat{\mathbf{C}}$

sauruses are associated with the objects in the environment that are encoded in the neural networks in the initial development stage from the birth to 3 years. Individual languages are associated with not only the objects in the environment but also the symbols that have already been incorporated in the environment. The same is true for cultural languages. In the Structured Meme Theory (SMT) [11], which is the memory construct to be used and updated by the MHP/RT, the patterns that represent the thesauruses, individual languages, and cultural languages are called action-level meme (A-meme), behavior-level meme (B-meme), and culture-level meme (C-meme), respectively.

As shown in the rounded corner rectangle "Memory" in Figure 2, the relationships between the three levels of memes and the Multi-Dimensional Memory Frame are as follows:

- A-memes represent bodily actions stored in the Motor-Multi-Dimensional Memory Frame.
- B-memes represent behaviors in the environment stored in the Behavior-Multi-Dimensional Memory Frame.
- C-memes represent culture stored in the Relation- and Word-Multi-Dimensional Memory Frame.

III. Details of Interaction between Self and O

This section describes how individual human perceptual, cognitive, and motor processes and memory work for each of the interaction types introduced in Section II, and clarifies how the symbolic grounding problem is involved and solved.

A. Interaction between Self and $\hat{\mathbf{O}}_{\mathbf{S}}$

The process of generating an object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ that represents the concept $\hat{\mathbf{C}}$ by self ($\hat{G}_{\hat{O}(\hat{C})}$) and perceiving the generated result ($\hat{P}_{\hat{O}(\hat{C})}$) is shown in Figure 3 using Figure 2 as a basis. According to this process, an object that matches the concept is generated. Figure 4 shows the flow of activation within the Multi-Dimensional Memory Frame involved in this process.

1) The concept $\hat{\mathbf{C}}$ is represented as a symbol within C-memes, and the nodes associated with it in the Word-Multi-Dimensional Memory Frame are activated.
2) Activation spreads from the Word- to Perceptual-Multi-Dimensional Memory Frame.
3) Activation spreads from the Perceptual- to Relation-Multi-Dimensional Memory Frame that encodes C-memes, and reaches Motor-Multi-Dimensional Memory Frame encoding A-memes via Behavior-Multi-Dimensional Memory Frame encoding B-memes.
4) According to the contents of activated Motor-Multi-Dimensional Memory Frame, the motor process operates to express bodily movements, which result in the generation of object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ in the environment.

Figure 4. Object Generation Process

This object generation process is indicated along the thick orange arrow in Figure 3. It is denoted symbolically as follows:

Generation Path [G-SS]

$$\hat{\mathbf{C}} \Rightarrow \left[\frac{\text{W-, R-, B-MDMF} \rightleftharpoons \text{M-MDMF}}{\text{P-MDMF}} \right] \Rightarrow \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \quad (1)$$

Here, the part indicated by [···] in the middle of the expression shows the process of spreading activation in the Multi-Dimensional Memory Frame from the viewpoint of the activated location in the memory. The symbol [G-SS] reads as follows; [**G**enerate - symbol in **S**elf via memory of **S**elf].

The generation process [G-SS] can be rearranged from the perspective of memes, i.e., memories of contents, as follows:

Meme-Mapping

$$\hat{\mathbf{C}} \Rightarrow \left[\frac{\text{C-memes} \rightleftharpoons \text{B-memes}}{\text{P-MDMF}} \rightleftharpoons \text{A-memes} \right]_{\text{M-S}} \Rightarrow \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \quad (2)$$

The generated object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ shown on the right end is associated with a state in which the C-, B-, and A-memes activated in the process of spreading activation in the Multi-Dimensional Memory Frame starting from the concept $\hat{\mathbf{C}}$ shown on the left end. The association is represented in the middle, $[\cdots]_{\text{M-S}}$, called "meme-mapping" from $\hat{\mathbf{C}}$ to $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$.

1) Perceive $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ and the activation spreads within the Perceptual-Multi-Dimensional Memory Frame.
2) The activation spread from the Perceptual- to Word-Multi-Dimensional Memory Frame results in the activation of symbol related to the perceptual representation.

Figure 5. Object Recognition Process

Figure 5 shows the object recognition process. The thick blue arrow in Figure 3 indicates this process. [R-SS] reads [**R**ecognize object - generated by **S**elf using memory of **S**elf].

Recognition Path [R-SS]

$$\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}) \Rightarrow \left[\frac{\text{P-MDFM}}{\text{W-MDMF}}\right] \Rightarrow \hat{\mathbf{C}} \tag{3}$$

In [R-SS], if the concept $\hat{\mathbf{C}}$ is strongly activated, then $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ correctly realizes $\hat{\mathbf{C}}$ in the real world. In this case, [G-SS] and [R-SS] are connected and closed, and the symbol $\hat{\mathbf{C}}$ and $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ are cognitively replaceable, which is represented by $\hat{\mathbf{C}} \equiv \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$. This state can be regarded as a state in which symbol grounding has been achieved within self (see Figure 6).

- Symbol $\mathbf{C}$ in C-memes activates B- and A-memes via the Perceptual-Multi-Dimensional Memory Frame to generate $\hat{\mathbf{O}}_{\mathbf{S}}(\mathbf{C})$.
- Perception of $\hat{\mathbf{O}}_{\mathbf{S}}(\mathbf{C})$ activates $\mathbf{C}$ and its associated activation pattern in the Multi-Dimensional Memory Frame.
- The perceptual representation of $\hat{\mathbf{O}}_{\mathbf{S}}(\mathbf{C})$ in the Perceptual-Multi-Dimensional Memory Frame is associated with $\mathbf{C}$ in Word-Multi-Dimensional Memory Frame.
- In the future, $\mathbf{C}$ in the Word-Multi-Dimensional Memory Frame activates perceptual representation of $\hat{\mathbf{O}}_{\mathbf{S}}(\mathbf{C})$ even if it does not exist in the real world, which enables the self to perceptually simulate the concept along with the activation of the Multi-Dimensional Memory Frame necessary to actually generate the object.

Figure 6. Symbol Grounding of Concept $\mathbf{C}$ in Self

B. The Symbol Grounding Problem in AI

The symbol grounding problem in AI shown in Figure 1 is solved by the fact that an activation pattern equivalent to the activation pattern of memes in the Multi-Dimensional Memory Frame that occurs in the process of generating self's $\hat{\mathbf{O}}_{\mathbf{S}}$ also occurs in AI. By ensuring that the meme-mappings occur within AI and self are equivalent, i.e., $[\cdots]_{\text{M-AI}} \equiv [\cdots]_{\text{M-S}}$, AI can be a substitute for self. This is summarized in Figure 7, where "Perceptual-Information-Encoding-in-AI" is the substitute for the Perceptual-Multi-Dimensional Memory Frame of human to perform A/D transformation to input the real world data to the AI system. Since the memes are knowledge passed down from generation to generation, it is considered possible to represent them by symbols. The Perceptual-Information-Encoding-in-AI can also be represented in symbols by encoding environmental information by sensors that perform the same function as sensory organs. The symbol grounding problem in AI is thought to be solved by elucidating memes.

1) The symbol $\hat{\mathbf{C}}$, which is common to self's, in C-memes activates the Perceptual-Information-Encoding-in-AI as well as the associated C-memes in AI.
2) B-memes are activated via the activated portion of Perceptual-Information-Encoding-in-AI.
3) The part of A-memes that overlap the activated B-memes is activated.

The steps 1, 2, and 3 constitute the meme-mapping of AI.

4) What is expressed by the activated A-memes is implemented in the real world via appropriate actuators.
5) Upon input of the object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ in AI, activation spreads in the Perceptual-Information-Encoding-in-AI, followed by the activation of the symbol $\hat{\mathbf{C}}$ in C-memes.
6) The part in the Perceptual-Information-Encoding-in-AI that corresponds to $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ and the symbol $\hat{\mathbf{C}}$ integrate the C-, B-, and A-memes activated in the steps 1, 2, and 3 to form an integrated association. At this point, $\hat{\mathbf{C}} \equiv \hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}})$ is established by AI by means of $[\cdots]_{\text{M-AI}}$, in other words, the symbol $\hat{\mathbf{C}}$ both the AI and the self commonly recognize has been grounded, secured by the relationship $[\cdots]_{\text{M-AI}} \equiv [\cdots]_{\text{M-S}}$.

Figure 7. Symbol Grounding of Concept $\hat{\mathbf{C}}$ in AI

C. Interaction between Self and $\bar{\mathbf{O}}_{\mathbf{R}}$

Consider the case where the other human generates an object. The object generation process for the concept $\bar{\mathbf{C}}$ performed by the other human is represented as $\bar{\mathbf{C}} \Rightarrow [\cdots] \Rightarrow \bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}})$, which is a the-other-human's version of [G-SS]. The other human spreads activation in his/her own Multi-Dimensional Memory Frame. The meme-mapping used for the other human is denoted as $[\cdots]_{\text{M-O}}$. The self interacting with the other human who has just generated $\bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}})$ recognizes it according to the following path:

Recognition Path for the Object Generated by Other [R-OS]

$$\bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}}) \Rightarrow \left[\frac{\text{P-MDFM}}{\text{W-MDMF}}\right] \Rightarrow \hat{\mathbf{C}} \tag{4}$$

If the symbol $\bar{\mathbf{C}}$ ($\equiv \bar{\mathbf{O}}_{\mathbf{R}}(\bar{\mathbf{C}})$) held by the other human and the symbol $\hat{\mathbf{C}}$ held by the self are identical, the symbol is transmitted through the object expressed by the other. For example, the other person holds a certain word $\bar{\mathbf{C}}$ in his/her mind and expresses it physically through gestures, and then the self sees it and assigns the word $\hat{\mathbf{C}}$ to it. The latent word of the other is connected to the self's latent word through the physical actions of the other person. Consider the case of communication via words, where the self and the other look at a sequence of words $\tilde{\mathbf{C}}$. The self and the other perform symbol grounding according to their respective

generation paths; $[\ \cdots\ ]_{\text{M-S}}$ and $[\ \cdots\ ]_{\text{M-O}}$ are included in each symbol grounding process. If the self and the other have grown up in the same environment, which is the necessary condition for them to have a common set of memes, then the relation $[\cdots]_{\text{M-S}} \equiv [\cdots]_{\text{M-O}}$ would hold, and the shared symbols have the same meaning. However, in the case of $[\ \cdots\ ]_{\text{M-S}} \neq [\ \cdots\ ]_{\text{M-O}}$, the meaning of all visually shared symbols may not be shared. For example, the phrase "see you on the ground floor" may trigger different behaviors depending on the culture to which the reader of the phrase belongs.

D. Interaction between Self and $\bar{\mathbf{O}}_{\mathbf{V}}$

In the case of interaction between self and other humans, the interaction is symmetric because both parties are humans. That is, in the part indicated by $[\cdots]$ that connects the symbol and the object in the generation and recognition paths, the activation spreads inside the Multi-Dimensional Memory Frame owned by self and others, respectively. On the other hand, in the case of Self–VR interaction, the generation and recognition paths on the system side are different from those on the human side. In the generation path, symbols defined in the system are converted into objects that can be perceived by the user. In the recognition path, human-generated objects are input to the system via sensors and converted into symbols that the system can handle. Both conversions are performed by programs implemented in the system.

In a VR environment, the system takes in the information emitted by human users and then determines the response to it. In any case, the input is represented as a symbol $\tilde{\mathbf{C}}$. Within the system, after setting a symbol $\bar{\bar{\mathbf{C}}}$ to be transmitted in the next cycle of interaction, the symbol-object transformation is performed and the object $\bar{\bar{\mathbf{O}}}_{\text{V}}(\bar{\bar{\mathbf{C}}})$ is output to be perceived by the user. This generation path is denoted as [G-VV].

The user perceives the object $\bar{\bar{\mathbf{O}}}_{\text{V}}(\bar{\bar{\mathbf{C}}})$ and recognizes it as a symbol along the recognition path [R-VS]. Let $\hat{\mathbf{C}}'$ be the recognized symbol. The user activates his/her Multi-Dimensional Memory Frame along the generation path [G-SS] for $\hat{\mathbf{C}}'$ and obtains the corresponding object $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}')$. If the relation $\hat{\mathbf{O}}_{\mathbf{S}}(\hat{\mathbf{C}}') \equiv (\text{or} \approx)\ \bar{\bar{\mathbf{O}}}_{\text{V}}(\bar{\bar{\mathbf{C}}})$ holds, the interaction will proceed smoothly. If not, it will fail.

IV. Method of $\bar{\mathbf{O}}_{\mathbf{V}}$ Generation with Symbol Grounding Secured

Only if the object generation path [G-VV] in the system is executed according to the user's meme mapping $[\cdots]_{\text{M-S}}$, i.e., if the meme mapping in the system is based on $[\cdots]_{\text{M-AI}}$, it is possible to proceed with an interaction that guarantees symbol grounding. Since memes are knowledge that are passed down from generation to generation, they can be represented by symbols. In this section, we add explanations for memes and suggest a method for externalizing them.

A. Getting Memes into the Brain

1) Action-Level Memes: During the period from birth to two to three years of age, humans generate a large number of synapses that connect neural circuits in the brain and take in as much information as possible. The rate of synapse generation is then reduced, and the distribution of information up to that point is used to determine the basic characteristics of the sensory organs. At the same time, by initiating body movements and imitating the movements of the people around them, they acquire body movements that have been formed empirically and accumulatively as individual ecology. This is formed through life's skillful method of adjusting the growth of muscles and other parts of the body to external constraints. At the same time, information from the sensory organs is linked to bodily actions. The most important bodily functions formed at this stage are the voice and hand functions.

2) Behavior-Level Memes: Later, the voice paves the way to speech; the hands pave the way to tool use. Through continuous imitation, humans learn to use the words and tools of those around them. At this time, humans acquire a new hierarchy of actions by organizing and summarizing the fact that a particular collection of sounds evokes a particular response, and that the feel of a hand experienced through tactile sensation and the movement of a tool perceived visually are captured as an unified whole via the tool. This is made possible by linking the A-memes formed on the brain circuit according to the situation in which they are used, and making them available as a coherent whole.

3) Culture-Level Memes: Furthermore, words pave the way to language, and tools pave the way to the use of more complex machines. At this stage, humans learn to act as members of the culture and civilization of the group to which they belong, not only through imitation but also through the experience of autonomous activities as members of the group. At this time, the B-level memes is extended to be used in a complex manner, and culture-specific behavior patterns are formed.

B. Mapping Memes into Information Systems

The mechanism by which the memes inherit information is analogous to an information system. A-memes serve as the operating system that defines general patterns of spatial-temporal behavioral functions. B-memes represent middleware that extends the general patterns to concrete patterns. C-memes act as application tools that extend the concrete patterns to the ones that work in a number of groups of people. By viewing memes as information systems, it is possible to represent human activities in various situations with symbols by observing them. Meme extraction has been attempted in some studies; the extraction of memes by observing the behavior of people trying to reach their destinations while acquiring information from information displays at railroad stations [5]; the inheritance of skills by ceramic artists through the acquisition of memes [14]; the memes used by skilled piano players during practice for a concert [15]. It is expected that memes in interaction in VR environments can also be elucidated based on these methods with appropriate modifications.

V. Conclusion

We argued that symbol grounding is realized when we perceive and recognize an object generated from a certain

symbol and match it with that symbol. In the object generation process, A-, B-, and C-memes, are involved in the conversion from symbols to objects. Since memes are passed down from generation to generation, we suggested that they could be extracted by observation, referring to the previous studies [5][14][15]. These are based on the MHP/RT [5][6][7] and the SMT [11][13]. The ability to evaluate objects that are presented when users interact with VR environments from the symbol grounding perspective is an important issue in the development of seamless VR environments.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334 , and National University Management Reform Promotion Project.

REFERENCES

[1] A. Newell, “Physical Symbol Systems,” Cognitive Science, vol. 4, no. 2, 1980, pp. 135–183.
[2] A. Newell, Unified Theories of Cognition (The William James Lectures, 1987). Cambridge, MA: Harvard University Press, 1990.
[3] J. E. Laird, A. Newell, and P. S. Rosenbloom, “Soar: An architecture for general intelligence,” Artificial Intelligence, vol. 33, 1987, pp. 1–64.
[4] “The symbol grounding problem,” Physica D: Nonlinear Phenomena, vol. 42, no. 1, 1990, pp. 335–346, [retrieved: 2, 2025].
[5] M. Kitajima and M. Toyota, “Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT),” Behaviour & Information Technology, vol. 31, no. 1, 2012, pp. 41–58.
[6] M. Kitajima and M. Toyota, “Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT),” Biologically Inspired Cognitive Architectures, vol. 5, 2013, pp. 82–93.
[7] M. Kitajima, Memory and Action Selection in Human-Machine Interaction. Wiley-ISTE, 2016.
[8] M. Kitajima et al., “Basic Senses and Their Implications for Immersive Virtual Reality Design,” in AIVR 2024 : The First International Conference on Artificial Intelligence and Immersive Virtual Reality, 2024, pp. 31–38.
[9] D. Kahneman, “A Perspective on Judgment and Choice,” American Psychologist, vol. 58, no. 9, 2003, pp. 697–720.
[10] D. Kahneman, Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux, 2011.
[11] M. Toyota, M. Kitajima, and H. Shimada, “Structured Meme Theory: How is informational inheritance maintained?” in Proceedings of the 30th Annual Conference of the Cognitive Science Society, 2008, p. 2288.
[12] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. W W Norton & Co Inc, 2 2018.
[13] M. Kitajima, M. Toyota, and J. Dinet, “How Resonance Works for Development and Propagation of Memes,” International Journal on Advances in Systems and Measurements, vol. 14, 2021, pp. 148–161.
[14] M. Kitajima, M. Toyota, and J. Dinet, “Art and Brain with Kazuo Takiguchi,” in COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, 2023, pp. 1–10.
[15] K. T. Nakahira, M. Kitajima, and M. Toyota, “Understanding Practice Stages for a Proficient Piano Player to Complete a Piece: Focusing on the Interplay Between Conscious and Unconscious Processes,” International Journal on Advances in Life Sciences, vol. 16, 2024, pp. 164–177.