Making VR Environments Fully Resonate with Interacting Humans: Core Constructs

Muneo Kitajima [1] (Presenter) Makoto Toyota [2] Katsuko T. Nakahira [1]

[1]Nagaoka University of Technology, JAPAN
mkitajima@kjs.nagaokaut.ac.jp

[2]T-Method, JAPAN

Muneo Kitajima is currently a professor emeritus at Nagaoka University of Technology. His recently published book ”Memory and Action Selection in Human-Machine Interaction” (2016) proposes a unified theory of action selection and development by integrating PDP, Two Minds, and layered structure of human action. The theory provides a comprehensive view of how our brain functionally works in our daily life. His current interest is to understand the implications of the theory to development of skill of adaptive problem solving, the important skill for survival.

Project HP: http://oberon.nagaokaut.ac.jp/ktjm/organic-self-consistent-field-theory/index.html
HP: http://oberon.nagaokaut.ac.jp/ktjm/index.html
email: mailto:mkitajima@kjs.nagaokaut.ac.jp

Abstract

- Much of human life is shaped by interacting with artificial objects in the surrounding environment. When these artificial objects behave not according to natural laws but according to algorithms created by artifact designers, the environment can be considered a virtual environment.
- Our previous research has conceptualized the human–object interaction process as comprising perceptual, cognitive, and motor processes operating synchronously with the environment's changes, alongside memory processes operating asynchronously with the environment's changes, further linked by resonance between them.
- More specifically, P-resonance triggered by the basic senses comprised of rhythmic, spatial, and number senses connects perceptual process with the P-MDMF, and C-resonance connects the MDMF with cognitive process via the structure of Goals, Operators, Methods, and Selection rules (GOMS) established in the MDMF.
- This paper presents the characteristics objects in a virtual environment should possess to enable smooth human–object interactions.

Humans acquire information about the external world through their five senses, activate Perceptual, Cognitive, and Motor (PCM) processes, and select and execute appropriate actions based on the circumstances at that moment. These processes are controlled by the brain, which forms a complex mechanism composed of a huge number of cells. However, when viewed functionally as an information processing device that appropriately converts input to the brain into output to the outside world, it can be understood as a very simple cognitive architecture model.

Classical Cognitive Architecture Models

Model Human Processor (MHP) [1]

- Card, Moran, and Newell [1] devised Model Human Processor (MHP), a cognitive architecture model that simulates human perception, cognition, and motor processes.
- By analogy to von Neumann-type computers as information processing devices, they conceived the brain as comprising the components “memory,” “processor,” and “program,” which defines how memory and processor are coupled.
  - *Memory is characterized by its storage capacity, decay time, and data type, consisting of visual and auditory image stores, working memory, and long-term memory;*
  - *Processor is characterized by its cycle time, which is the processing time per unit task.*
- The target domain involves solving problems within a clearly defined problem space based on the principle of rationality to achieve the specified goal.
- The knowledge required in this situation is organized as Goals, Operators, Methods, and Selection rules (GOMS).

Soar [2, 3]

- By encoding perceived information within the perceptual process and representing it as symbols, it becomes possible to *think* within the cognitive process.
- In the thinking process, memory is utilized, and symbols are successively transformed. Memory itself is also expressed through symbols. Part of the result of thinking provides an expression of a sequence of actions executable in the motor process.
- Based on the idea that human intelligence can be captured by the thinking process of manipulating symbols, Newell proposed the Physical Symbol System as a theory of human intelligence [4].
- This approach provided the foundation for Soar, a cognitive architecture that views thinking as problem-solving process involving manipulation of a problem space [2, 3].

However, the real-world environment with which humans interact changes its state moment by moment based on the mechanisms inherent within the environment itself. Mechanisms can be either linear or nonlinear. In the latter case, it is fundamentally impossible to predict the temporal evolution of the state. When confronted with such real-world environments, it is difficult to fully capture the appropriate actions required in response to the environment's state as a rational problem-solving process within a predefined, well-defined problem space.

MODEL HUMAN PROCESSOR WITH REALTIME CONSTRAINTS (MHP/RT)

Outline of MHP/RT [5, 6]

- We have developed MHP/RT, that can simulate not only behaviors modeled by navigation within well-defined problem spaces targeted by MHP and Soar, but also adaptive and flexible behaviors. MHP/RT consists of the following three elements:

1) "Perception-Cognition-Motor (PCM) Process," which selects and executes actions that are likely to yield satisfactory results without causing the human system to fail. This process *synchronizes* human interactions with the real-world environment along the time axis with the state transitions of that environment;
2) "Memory Process," which operates *asynchronously* with the real environment, accumulating the results of PCM process execution and functioning to be utilized during action selection and execution;
3) "Resonance" mechanism connecting the *environment-synchronized* PCM process and the *environment-asynchronous* memory process.

Understanding Humans Acting in Environment

- The key issues are as follows.

1) PCM processes operating in various modes depending on the task;
2) The structure and content of memory acquired while PCM processes are active and utilized by them;
3) P-resonance linking perceptual processes and memory;
4) C-resonance linking memory and cognitive processes.

- Understanding is achieved by holistically relating these elements and grasping their overall dynamics.

This paper identifies factors to consider when analyzing and understanding them, structuring insights gained during our research process according to the four points listed above. From the perspective of ensuring smooth interaction with the environment, it outlines the necessary requirements for achieving it.

This section provides an overview of the PCM process and memory process within MHP/RT, a cognitive architecture capable of simulating everyday behavioral selections [7, 5, 6] as the foundation.

Human–Object Interaction

- Interaction between humans and the object $\hat{O} \in \mathbf{Environment}$ in the environment is realized through PCM processes operated by humans, memory processes utilized by PCM processes and updated as a result of PCM process execution, and processes occurring within $\hat{O}$.
- $\hat{O}$ changes state over time. These changes may arise either from human actions affecting $\hat{O}$ or from mechanisms inherent within $\hat{O}$ itself.
- In either case, for humans to smoothly continue their interaction with $\hat{O}$, it is necessary to synchronize their own PCM process with changes in the state of $\hat{O}$.

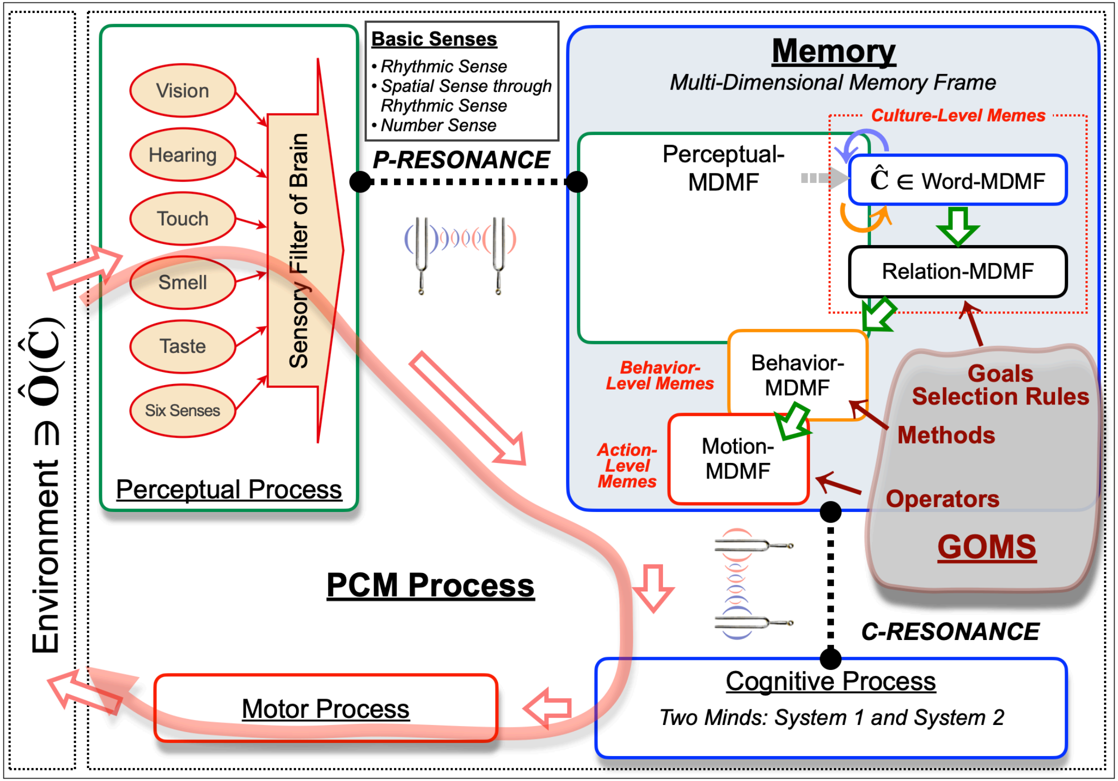
Figure on the next slide shows its overall outline.

PCM and Memory Processes, and Resonance

- Meanwhile, memory defines the content executed by the PCM process. Memory accumulates the results of the PCM process's execution. Furthermore, memory is provided during the execution of the PCM process. The processes of accumulation and provision are executed asynchronously with respect to the environment's temporal development.
- PCM processes operating synchronously with the environment and memory processes operating asynchronously with the environment collaborate to realize human behavior within the environment—what is executed and how—by linking them through resonance.
- Resonance includes P-resonance occurring between perceptual processes and memory, and C-resonance occurring between cognitive processes and memory.

These enable the PCM process to utilize memory and generate behavior within the environment.

PCM processes operating *synchronously* with the environment and Multi-Dimensional Memory Frame operating *asynchronously* with the environment



created by combining [8, Figure 2] and [9, Figure 6(b)]

PCM and Memory Processes, and Resonance

- *PCM Process:* The red arrows in Figure on slide 6 show the process by which environmental information is incorporated into the body via sensory nerves, undergoes processing within the brain, and then acts upon the external world via motor nerves, based on MHP/RT [6, 5].
- *Memory:* The MDMF comprises the P-, B-, M-, R-, and W-MDMF. The P-MDMF overlaps with the B-, R-, and W-MDMF. This allows activity to propagate from P- to M-MDMF.
- *P-Resonance:* Perceptual information taken in from the environment through sensory organs *resonates* with information in the MDMF, which is called P-Resonance [10].
- *C-Resonance:* System 2 utilizes the W- and R-MDMF via C-Resonance, while System 1 draws on the B- and M-MDMF via the same mechanism. Motor sequences are then expressed according to the M-MDMF.
- *Memory Update:* 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.

Four Operation Modes of MHP/RT

- MHP/RT's action selection process is controlled by System 1 and System 2. These systems cooperate to link perception and movement, and the degree of cooperation depends on the state of the external environment with which the MHP/RT interacts.
- The interaction is carried out in one of the four operation modes.
- Interactions in Mode 1 or 2 are sound in the sense that System 2 monitors whether actions executed by System 1 in a timely manner have deviated from the desired trajectory.
- Meanwhile, Mode 3 and 4 contain unmonitored feedforward System 1 processes.
- In Mode 2, System 2 frequently intervenes the PCM processes conducted by System 1.
- In Mode 1, System 2's intervention is weak. The external environment supports the automatic processes performed by System 1, enabling smooth interaction. *This is the smooth interaction addressed in this paper.*

This study proposes that the smooth interaction is achieved through the occurrence of P-resonance and C-resonance.

When the PCM process is running, the contents of P-MDMF are updated in response to the perceptual process, those of W-, R-, and B-MDMF are updated in response to the cognitive process, and those of M-MDMF are updated in response to the motor process.

..

Structure of Memory

- The memory system—Multi-dimensional Memory Frame—shown in the upper right of figure on slide 6 serves as traces of PCM process operation. It is classified into Perceptual-, Word-, Relation-, Behavior-, and Motor-Multi-dimensional Memory Frame.

Accumulating Contents through Imitation

- Imitation—doing what one observes—forms the basis of behavior. Therefore, the results of imitative behavior are structured and represented within the MDMF, and imitative behavior arises through the propagation of activity within the MDMF.
- Imitation occurs between individuals, and imitative behaviors exist that are passed down across generations. Therefore, the MDMF can be organized from the perspective of *memes* inherited across generations [11].

Structured Meme

- Words, the archetype of Meme [12], exist within the individual's W-MDMF as language. Through the use of language, individual languages or cultural languages are organized within the MDMF.
- Individual language—*Behavioral-level Memes, stored in the B-MDMF*—is the language used in person-to-person communication, encompassing not only direct but also metaphorical usage.
- Cultural language—*Cultural-level Memes, stored in the R- and W-MDMF*—is language used within a cultural context where a proper understanding of the established common sense within a specific community is essential for successful communication.
- These languages develop during the early developmental stage from birth to age three, associated with objects in the environment —*Motor-level Memes, stored in the M-MDMF*— encoded within neural networks.

These memes circulate among people and persist across generations [13].

To generate appropriate actions synchronized with the environment in ever-changing situations, it is necessary to bind spatiotemporal information to MDMF and render M-MDMF executable. As a foundational concept to solve this binding problem, the concept of basic senses was proposed, which was identified as the actual manifestation of P-resonance [10]. P-Resonance occurs between the external stimulus and the P-MDMF by rhythmic and spatial senses. Subsequently, cognitive objects are generated by utilizing the number sense.

Rhythmic Sense

The changes brought about by human action alter the relative situation between the actors themselves and the environment.

- *Human*: The human organs have evolved and developed under circadian rhythms. Consequently, periodically active organs such as the heart have been formed to provide unique rhythms.
- *Environment*: The changes with various reproducible rhythms occur under the cyclic activity of the earth.

To adapt to changes in the environment, *rhythmic sense* should be formed. The rhythmic sense enables flexible binding of memory and perceptual information on the time axis in P-Resonance.

Spatial Sense

Bodily activity includes movement that involves changing the position of one's own body part in the 3D space.

- Recognition of the current situation is necessary for constructing executable activities from information stored in the M-MDMF, free from absolute positions.
- The unique dimensions associated with movement are distance and time.
- The time is associated with body's internal rhythms, which define the scale for measuring distance.

The information concerning distance between objects is conceived through the rhythm-based scale, which is *spatial sense* defined through the rhythmic sense.

Number Sense

For humans to select appropriate actions in a timely manner, information related to quantitative comparisons is indispensable.

- When this information is combined with the reward response, a basic sense of quantitative discrimination is formed, which is *number sense* [14].
- Perceptual information activates P-, R-, and B-MDMF through P-resonance mediated by rhythmic and spatial senses.
- Number sense integrates them into cognitive objects that can be consciously manipulated.

The number sense activates these regions, making the MDMF available for cognitive processes via C-resonance.

C-resonance occurs between cognitive objects represented as activation patterns in P-, B-, R-, and W-MDMF and System 1 and System 2, driving the cognitive process. As a result, activation propagates to M-MDMF, mapping to the brain's representation of motor movement. The MDMF contains three hierarchically structured memes—C-, B-, and A-memes. These memes correspond to each memory in the MDMF. Associations between memes form a structure that binds them together. This ensures the reality derived from perceptual information linked to the real world.

Implementation of C-resonance

C-resonance operates under time constraints where PCM process must synchronize with environmental changes while selecting and executing actions, linking the MDMF with cognitive processes.

- Our previous work introduced GOMS[1] as a mechanism for implementing C-resonance to directly link the W-, R-, B-, and M-MDMF without going through P-MDMF [9].
- Using memes in an interpreter-like manner via P-MDMF ensures corporeality; it requires monitoring while advancing the PCM process, making it inefficient.
- The use of MDMF in the form compiled by GOMS corresponds to situations where MHP/RT is running smoothly in Mode 1 under System 1 control.

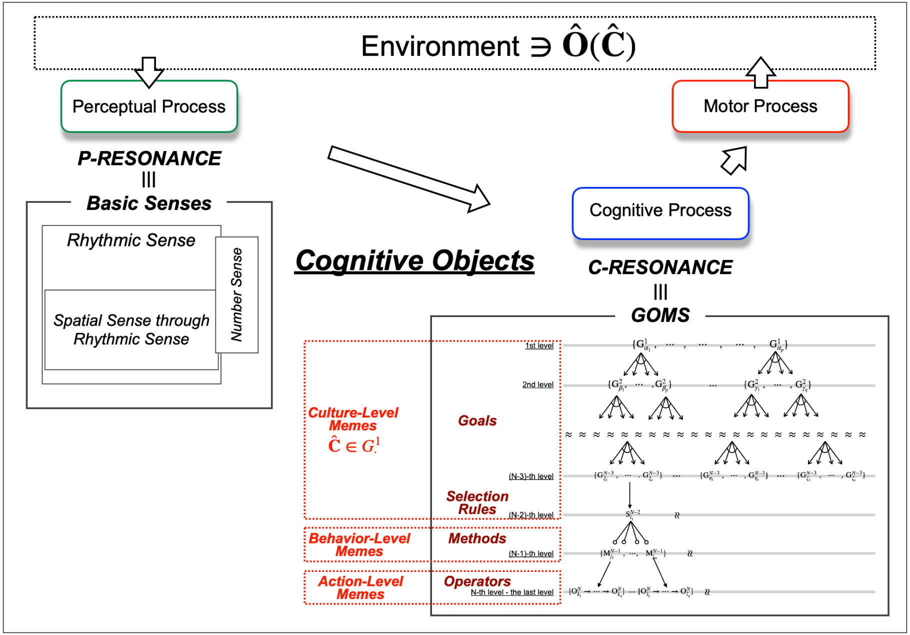
Binding Memes via GOMS

- In GOMS, behavioral goals form a robust hierarchical structure, which mediates the organization of behavior. This structure does not hold the time as its primary parameter.
- GOMS is used to structure A-, B-, and C-memes which do not contain absolute spatiotemporal information as a method of realizing behavior generation without breaking down, while keeping in sync with the real world where the situation changes from moment to moment.
- GOMS should correspond to the phenomenon of A-, B-, and C-memes binding without the P-MDMF.
- GOMS is a shortcut that guarantees symbol grounding without going through perception.

Symbol Grounding via GOMS

- For the concept or goal $\hat{C} \in$ C-memes that the self intends to realize in the real world, the self propagates activations through P- to R- and B-MDMF, ultimately activating M-MDMF. Then, the object $\hat{O}(\hat{C})$ embodies itself in the real world through the motor process.– *see slide 6*
- This object is perceived by the self, and the P-MDMF is activated via P-resonance. Then the activation propagates to W-MDMF.
- When the object is recognized as $\hat{C}$, and the relationship $\hat{C} \equiv \hat{O}(\hat{C})$ holds true, meaning the envisioned concept has been successfully realized in the real world.
- The establishment of this relationship also signifies the resolution of the symbol grounding problem [15].

Environment-synchronized PCM processes and activation of environment-asynchronous MDMF



This figure illustrates that the following occurs for the concept $\hat{C}$ ∈ the W-MDMF—C-memes—corresponding to the state one intends to achieve. First, the cognitive process resonates with GOMS, which includes the corresponding goal, activating System 2 and System 1, thereby generating the object $\hat{O}(\hat{C})$ in the environment. Subsequently, this object captures the basic senses, inducing P-resonance and forming cognitive objects in the MDMF. If the cognitive objects are linked to $\hat{O}$, symbol grounding is guaranteed. Next slide shows key design considerations to ensure sufficient C-resonance and P-resonance in artificial objects.

Promoting C-Resonance

Since the true nature of C-resonance is GOMS, it is necessary to represent situations where virtual environments are utilized using GOMS. This is achieved by constructing a GOMS representation of human behavioral ecology within the target environment.

- GOMS organizes memes into a hierarchical structure; the procedure for constructing GOMS begins with extracting memes. Here, a method known as Cognitive Chrono-Ethnography (CCE) [16] proves effective.
  - A-memes are observable. Mutually distinguishable ones define the operators "O".
  - B-memes are pattern sequences associated with methods "M", which are patterns reproducibly generated when conditions are met by sequences of A-memes.
  - The conditions define selection rules "S", expressed as "IF condition THEN pattern sequence name" processed by System 2.
  - Upon this, goal structures—C-memes—unfold. At the top, goals related to happiness and satisfaction are placed.

C-resonance is promoted by accurately grasping in advance the goal, identifying the goal being pursued from the observed sequence of actions, and appropriately inferring the next goal that will likely be set.

Promoting P-Resonance

The basic senses are applied to generate cognitive objects from the P-resonated portion of P-MDMF. Since $\hat{\mathrm{O}}$ serves as the starting point for P-resonance, it is necessary to design $\hat{\mathrm{O}}$ to elicit responses through basic senses.

- *Rhythmic Sense:* The most fundamental element of rhythm is the existence of regularity or repetitive patterns in the intervals between events. The events arising from interaction generating patterns along the time axis become a necessary condition for rhythm perception to be triggered.
- *Spatial Sense:* The determination of rhythm signifies that the unit time $\hat{T}$ is established; the difference between the adjacent events $D(E(T+\hat{T}), E(T))$ becomes discernible. Objects can evolve over time according to their own programmed dynamics and generate events; it is necessary to consider that the receiving human side quantizes events according to a rhythm characterized by the unit time $\hat{T}$.
- *Number Sense:* It is used to select information concerning quantitative comparisons of about three items. The evaluation can be conducted in a multifaceted manner, but comparisons involving three or fewer items are preferable.

What This Paper Has Done

- Interaction is executed through a PCM process that synchronizes with the ever-changing environment.
- The execution content is determined by the activation patterns of the MDMF, which operate independently of environmental changes.
- To maintain smooth interaction, it is important to control the content of P-resonance generated by basic senses and the content of C-resonance generated by GOMS. This paper has outlined the matters that should be considered in these controls.

Guideline Focusing on MHP/RT's Operation Modes

- Guidelines are generally expressed in abstract terms, so to apply them to real-world artificial environment design scenarios, implementation methods for the guidelines must accompany them.
- This study showed that designs aiming to induce P-resonance and C-resonance, and designs enabling the functioning of basic senses and GOMS, are effective as practical implementations of the guidelines [17] proposed in relation to the MHP/RT operating modes (specifically Mode 1).

Next Challenge

- Just as practice became possible by focusing on *basic senses* and *GOMS—concepts* external to MHP/RT—the next challenge is to make artifact design guidelines, derived from *weak synchronization (the conditions for P-resonance) associated with MHP/RT's four processing modes*, practicable by introducing concepts external to MHP/RT.
- Here, ***dynamic*** *analysis focusing on the nature of interactions over time is fundamental.* For example, the degree of weak synchronization in situations where latent information about visual objects presented in a VR learning environment is provided auditorily can be examined by measuring visual behavior and pupil responses, and analyzing the relationship between these results and memories of audiovisual experiences [18, 19].

[1] Stuart K. Card, Thomas P. Moran, and Allen Newell. ***The Psychology of Human-Computer Interaction.*** Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[2] Allen Newell. ***Unified Theories of Cognition (The William James Lectures, 1987).*** Harvard University Press, Cambridge, MA, 1990.

[3] Johnson E. Laird, Allen Newell, and Paul S. Rosenbloom. **Soar: An architecture for general intelligence.** *Artificial Intelligence*, 33:1–64, 1987.

[4] Allen Newell. **Physical Symbol Systems.** *Cognitive Science*, 4(2):135–183, 1980.

[5] Muneo Kitajima and Makoto 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*, 5:82–93, 2013.

[6] Muneo Kitajima. ***Memory and Action Selection in Human-Machine Interaction.*** Wiley-ISTE, 2016.

[7] Muneo Kitajima and Makoto Toyota. **Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT).** *Behaviour & Information Technology*, 31(1):41–58, 2012.

[8] Muneo Kitajima, Makoto Toyota, and Katsuko T. Nakahira. **Why the Symbol Grounding Problem Matters in Virtual Reality: A Meme-Focused Solution Based on the Model Human Processor with Real-Time Constraints.** *International Journal On Advances in Intelligent Systems*, 18(3 & 4):162–172, 2025.

[9] Muneo Kitajima, Makoto Toyota, Jérôme Dinet, and Katsuko T. Nakahira. **Transforming Conscious Goals into Unconscious Actions in Real-world Interactions: Real-world Use of Behavioral Ecological Memes via GOMS.** *International Journal On Advances in Intelligent Systems*, 18(3 & 4):173–186, 2025.

[10] Muneo Kitajima, Makoto Toyota, Jérôme Dinet, Anissa Bystrzycki, Anne-Marie Eby, and Marie Gérard. **Basic Senses and Their Implications for Immersive Virtual Reality Design.** In *AIVR 2024 : The First International Conference on Artificial Intelligence and Immersive Virtual Reality*, pages 31–38, 2024.

[11] Makoto Toyota, Muneo Kitajima, and Hideaki Shimada. **Structured Meme Theory: How is informational inheritance maintained?** In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, page 2288, 2008.

[12] Daniel C. Dennett. ***From Bacteria to Bach and Back: The Evolution of Minds.*** W W Norton & Co Inc, 2 2018.

[13] Muneo Kitajima, Makoto Toyota, and Jérôme Dinet. **How Resonance Works for Development and Propagation of Memes.** *International Journal On Advances in Systems and Measurements*, 14:148–161, 2021.

[14] Stanislas Dehaene. ***The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition (English Edition).*** Oxford University Press, 4 2011.

[15] Stevan Harnad. **The symbol grounding problem.** *Physica D: Nonlinear Phenomena*, 42(1):335–346, 1990.

[16] Muneo Kitajima. **Cognitive Chrono-Ethnography (CCE): A Behavioral Study Methodology Underpinned by the Cognitive Architecture, MHP/RT.** In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pages 55–56. Cognitive Science Society, 2019.

[17] Muneo Kitajima, Makoto Toyota, and Jérôme Dinet. **Guidelines for Designing Interactions Between Autonomous Artificial Systems and Human Beings to Achieve Sustainable Development Goals.** *International Journal on Advances in Intelligent Systems*, 15:188–200, 2022.

[18] Yuta Kurihara, Motoki Shino, Katsuko Nakahira, and Muneo Kitajima. **Visual behavior based on information foraging theory toward designing of auditory information.** In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: HUCAPP*, pages 530–537. INSTICC, SciTePress, 2024.

[19] Rinki Hirabayashi., Motoki Shino., Katsuko Nakahira T.., and Muneo Kitajima. **How auditory information presentation timings affect memory when watching omnidirectional movie with audio guide.** In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP,*, pages 162–169. INSTICC, SciTePress, 2020.