Multimodal Interactions Viewed as Dual Process on Multi-Dimensional Memory Frames under Weak Synchronization

Muneo Kitajima

Information and
Management Systems Engineering
Nagaoka University of Technology
1603-1 Kamitomioka,
Nagaoka, Niigata 940-2188, JAPAN
Email: mkitajima
@kjs.nagaokaut.ac.jp

Jérôme Dinet

Psychology and Neuroscience Laboratory
University of Lorraine
54021 Nancy Cedex, France
Email: jerome.dinet
@univ-lorraine.fr

Makoto Toyota

T-Method
Chiba Japan
Email: pubmtoyota@me.com

***Abstract*—Behavior of users interacting with multimodal interfaces looks complex because the degrees of freedom of sensory input and motor output is large. This paper suggests that this complexity can be alleviated by applying the Simon's ant metaphor to the multimodal interaction situations, i.e., "What users do is simple. 1) they use perceptual input to generate a mirror image of the real world surrounding the self to be shared in conscious and unconscious processes, 2) they select next actions by consciously planning ahead and unconsciously tuning motor movements for the event to happen, and after performing the action, they unconsciously modify the participated neural network and consciously reflect on the result of action, and 3) they perform 1 and 2 in synchronous with the ever-changing external environment, which this paper calls "weak synchronization." A cognitive architecture, Model Human Processor with Realtime Constraints (MHP/RT) and its associated memory structure, Multi-Dimensional Memory Frames, developed by the authors is briefly introduced considering the situation of users interacting with multimodal environment. Then, the above three items are derived as the essential principles for organizing user's behavior in multimodal interaction environment. Future work on designing mixed reality multimodal interaction environment is introduced that has its basis on the perspectives for multimodal interactions this paper claims.**

***Keywords*–Dual process; Multimodal interaction, Two Minds, Multi-dimensional memory frames; Weak synchronization; Time scale of human action.**

I. INTRODUCTION

Multimodal interaction provides multiple communication channels between users and automated systems. Speeches, finger movements, hand movements, eye movements, and so on, can be used for transmitting messages from users to systems. Synthesized voices, sounds, vibrations, texts and graphics rendered on visual display can be used for sending messages from systems to users. What happens at the multimodal interface between users and systems *looks vey complex* because the degrees of freedom of both sides is large. Multimodal interaction using virtual reality technologies to realize immersive environments is an emerging and continuously evolving domain, in which development of a theoretical framework to reduce the inherent complexity is required for advancing principled design of multimodal interaction [1].

This paper suggests an approach for alleviating this complexity by applying the metaphor of Simon's ant to the situation where users are engaging in multimodal interaction to use automated systems. Simon describes in the book entitled "The Sciences of Artificial" [2, pp.51–53] as follows:

> *An ant [A man], viewed as a behaving system, is quite simple. The apparent complexity of its behavior over time is largely a reflection of the complexity of the environment in which it finds itself.*

In these sentences, he described a situation where an ant produced a very complex path across the terrain of a beach while making decisions on which direction to go at each moment when it encounters an obstacle. A person observing only the path itself might be inclined to ascribe a great deal of intelligence of ant. However, it turned out that the complexity of the path is really produced by the complexity of the terrain over which the ant was navigating. The ant only selected the optimal operator from doable simple alternatives for the specific situation where it was needed to reduce the distance from the current location to its nest. The specific trajectory was the result of successive decisions of selecting locally optimal simple operators.

Simon [2] claims that decision making in problem solving situations, which looks complex and intractable, is in reality governed by two principles, i.e., *satsficing principle* and *bounded rationality*. The role of these behavioral principles for understanding a variety of decision making behaviors in problem solving situations is the same as that of the Newtonian equation of motion for predicting configurations of the planets of the solar system at specific times in the future. They should provide a firm basis for considering behaviors of decision makers; any models for explaining and predicting behaviors of decision makers in specific situations have to be constructed on them.

Users interacting with multimodal interfaces are placed in the situation where they have to select next actions and execute them by acting to the external environment through their body parts, e.g., limbs, eye balls, and so on, via motor neurons as serial processing, and receive signals from not only the external systems but also themselves (i.e., seeing their finger movements and hearing their voices) through the five senses, i.e., taste, sight, touch, smell, and sound, via sensory neurons as parallel processing. The received signals traverse the networks of intermediate neurons to select next actions. This is a circular

 ISBN: 978-1-61208-705-4

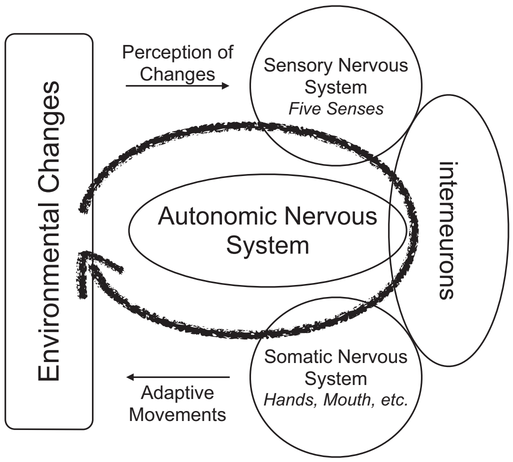

Figure 1. Continuous cyclic loop of perception and movement (adapted from [3, Figure 1]).

process as shown by Figure 1. After s/he perceives the results of movements of his/her body parts, as well as the changes of the external environment, a next Perceptual–Motor cycle occurs. Interneurons in-between the sensory neurons and motor neurons convert the input patterns to the output patterns – these constitute a Perceptual–Cognitive–Motor (PCM) process. Memories associated with the respective activities of sensory, motor, and intermediate neurons continuously accumulate and change as the PCM cycle runs. In multimodal interaction situations, the PCM cycle equipped with the memories associated with the respective processes runs in synchronous with the ever-changing multimodal environment, otherwise the user and the system are not able to establish "interaction." Given the very basic architecture of PCM cycle shown by Figure 1 and its associated memories, how synchronization between a user and a multimodal system could be established? This paper suggests a form of synchronization, "weak synchronization", should serve as the operational principle for the PCM architecture, as the satisficinig principle and bounded rationality do for decision making in a problem space.

This paper is organized as follows. Section II introduces a theory of action selection and memory that the authors have developed [4][5][6], which essentially elaborates the basic idea of endless cycle of PCM as depicted in Figure 1, and defines a cognitive architecture that would be most suitable for understanding people interacting with ever-changing environment. Section II-A introduces the Model Human Processor with Realtime Constraints (MHP/RT) that defines the PCM cycle, and Section II-B shows the memory system that accompanies with MHP/RT. Section III starts with explanations how MHP/RT interacts with multimodal environment, then, derives an operation principle "weak synchronization" as a critical means for a user as modeled by MHP/RT to interact smoothly with multimodal environment. Finally, Section IV provides a summary of the paper with future work focusing on how the insight of this paper could be used to better understand users interacting with multimodal environment.

II. Theory of Action Selection and Memory

Starting from the basic cycle of PCM processes depicted by Figure 1, Kitajima and Toyota [5][6] have constructed a comprehensive theory of action selection and memory, MHP/RT, that should provide a basis for constructing any models for users interacting with multimodal environment. The theory integrates the fundamental characteristics of human beings interacting with ever-changing environments. This section introduces briefly MHP/RT that defines the PCM processes and its associated memory, Multidimensional Memory Frames, that is used and modified while MHP/RT works.

A. MHP/RT

MHP/RT is an extension of Model Human Processor (MHP) developed by Card, Moran and Newell in 1983 [7]. MHP is a cognitive architecture to simulate users interacting with *then-available* information devices, such as The Star workstation, officially named Xerox 8010 Information System. MHP/RT aims at simulating users who interact with richer in contents and more dynamic modern information environment, such as multimodal interaction environment. MHP/RT has to deal with more information-rich situations than MHP was supposed to do.

MHP/RT implements at a higher level the following three facts concerning processing with an assumption. The facts are as follows:

1) The fundamental processing mechanism of brain is Parallel Distributed Processing (PDP) [8],
2) Human behavior emerges as the results of competition of the dual processes of System 2, a slow *conscious* process for deliberate reasoning with feedback control, and System 1, fast *unconscious* process for intuitive reaction with feedforward control for connecting perception and motor, called Two Minds [9][10], and
3) Behavior is organized under happiness goals [11], e.g., target happiness, competitive happiness, cooperative happiness, etc.

The assumption is that the endless PCM cycle continues from his or her birth to death in the ecological system, consisting of a person and his/her environments, as a periodic circulation system, called autopoiesis [12]. The system is truly dynamic and evolves in the irreversible time dimension.

In summary, MHP/RT essentially defines a specification for an organic version of PDP by incorporating Two Minds and happiness goals in the original version of PDP. The term, "Organic Parallel Distributed Processing (O-PDP)", was first introduced by [13]. O-PDP develops cross-networks of neurons in the brain as it accumulates experience of interactions in the environment. The neural network development process is circular, which means that any experience at a particular moment should reflect somehow the experience of the past interactions that have been recorded in the shape of current neural networks. In this way, an O-PDP system is organized evolutionally, and realized as a neural network system, including the brain, the spinal nerves, and the peripheral nerves to construct an O-PDP system.

MHP/RT, illustrated in the left portion of Figure 2, describes the cyclic PCM processes. It consists of four major autonomous systems: Perceptual Information Processing for perception, Autonomous Automatic Behavior Control Processing (System 1) and Conscious Information Processing (System 2) for cognition, and Behavioral Action Processing for motor

 ISBN: 978-1-61208-705-4

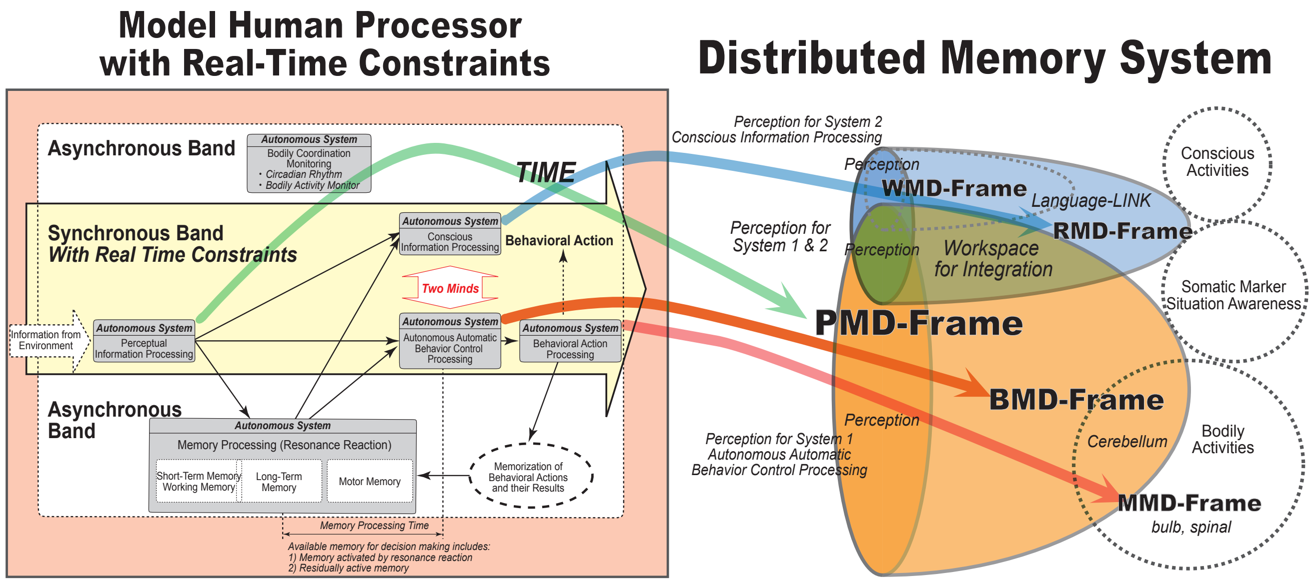

Figure 2. MHP/RT ([6, Figure 3]) and the distributed memory system.

movement. These processes work in synchronous with the ever-changing external environment, which effectively imposes real time constraints on the PCM processes. In addition, these processes connect with Memory Processing autonomous system to make use of stored contents in memories via resonance reaction, which happens not synchronously but asynchronously with the environmental changes.

The cyclic PCM processes are implemented in O-PDP as hierarchically organized bands having their respective characteristic times for operations, i.e., Biological, Cognitive, Rational, and Social bands defined as Newell's time scale of human action (Figure 3). Respective bands have their characteristic times. A number of phenomena that occur in a certain single band would be related with each other and therefore they could have linear relationships. On the other hand, they should have non-linear relationships with those phenomena that happen in a different band. For example, conscious activities in Rational Band, System 2, cannot have linear relationship with unconscious activities in Biological Band, System 1, but have non-linear relationships.

The existence of gaps between bands indicates that a phenomenon in one band evolves quasi-independently with another in a different band. However, one side would have some effect on the other in order to organize the activities of the O-PDP system coherently in the environment. System 1 runs quasi-independently with System 2 with occasional exchanges over the gap, e.g., conscious process, System 2, intervenes unconscious process, System 1, before unconscious feedforward process would derail. In many cases in our daily life, application of irrational cognitive bias could be avoided by deliberate thinking by System 2. This is a kind of synchronization between System 2 and System 1 when action selection is done by MHP/RT. This interaction is shown in the left portion of Figure 2 as “Two Minds” that connects Autonomous Automatic Behavior Control Processing (System 1) and Conscious Information Processing (System 2). As described, synchronization is an important mechanism to make the behavior of O-PDP system stable. This issue will be further discussed later in Section III-C.

MHP/RT is a *real* brain model comprising of System 1's unconscious processes and System 2's conscious processes at the same level as shown in Figure 2, in which both System 1 and System 2 receive input from the Perceptual Information Processing autonomous system in one way, and from the Memory Processing autonomous system in another way. System 1 and System 2 work autonomously without any superordinate-subordinate hierarchical relationships but interact with each other when necessary [6].

This feature of MHP/RT should be contrasted with the goal-oriented cognitive architectures such as ACT-R [14][15] in which the conscious processes are considered as the processes to control people's behavior and the unconscious processes are considered subordinate to the conscious or intentional processes [6]. What ACT-R tries to do is to show how System 2 can be implemented on top of System 1. The procedural memory system is very similar to System 1, and then ACT-R models tend to consist of a set of production rules that give rise to the slower, deliberative planning behaviors seen in System 2. This is a very different conceptualization for autonomously behaving creatures in the ever-changing environment from that given in this paper. However, ACT-R models are totally adequate for simulating stable human activities with weak time constraint in which deliberate decision making would work effectively, but might be hard for the situations with strong time constraint where the environmental condition changes chaotically and deliberate decision making implemented on System 2 might not work as effective. Multimodal interaction is one of those human activities that goal-oriented cognitive architectures would not be suitable.

B. Multidimensional Memory Frames

As illustrated in the right portion of Figure 2, each autonomous system for carrying out PCM processes in the synchronous band of MHP/RT is associated with its corresponding memory, which is implemented as a distributed memory

 ISBN: 978-1-61208-705-4

Time Sale of Human Action

Scale (sec)	Time Units	System	World (Theory)
10^{7}	months		Social Band
10^{6}	weeks		
10^{5}	days		
10^{4}	hours	Task	Rational Band
10^{3}	10 min	Task	
10^{2}	minutes	Task	
10^{1}	10 sec	Unit Task	Cognitive Band
10^{0}	1 sec	Operations	
10^{-1}	100 ms	Deliberate Act	
10^{-2}	10 ms	Neural Circuit	Biological Band
10^{-3}	1 ms	Neuron	
10^{-4}	100 µs	Organelle	

Figure 3. Newell's time scale of human action (adapted from [16]).

system. The contents of memory is structured as specified by the Structured Meme Theory [17]. A brief explanation of the respective multi-dimensional memory frames is as follows [18]:

- **PMD (Perceptual Multi-Dimensional)-frame** constitutes perceptual memory as a relational matrix structure. It collects information from external objects followed by separating it into a variety of perceptual information, and re-collects the same information in the other situations, accumulating the information from the objects via a variety of different processes. PMD-frame incrementally grows as it creates memory from the input information and matches it against the past memory in parallel.
- **MMD (Motion Multi-Dimensional)-frame** constitutes behavioral memory as a matrix structure. The behavioral action processing starts when unconscious autonomous behavior shows after one's birth. It gathers a variety of perceptual information as well to connect muscles with nerves using spinals as a reflection point. In accordance with one's physical growth, it widens the range of activities the behavioral action processing can cover autonomously.
- **BMD (Behavior Multi-Dimensional)-frame** is the memory structure associated with the autonomous automatic behavior control processing. It combines a set of MMD-frames into a manipulable unit.
- **RMD (Relation Multi-Dimensional)-frame** is the memory structure associated with the conscious information processing. It combines a set of BMD-frames into a manipulable unit.
- **WMD (Word Multi-Dimensional)-frame** is the memory structure for language. It is constructed on a very simple one-dimensional array.

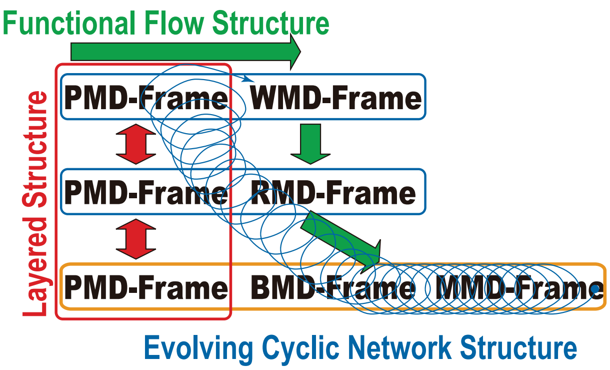

Figure 4. Multi-dimensional memory frames characterized by functional flow structure, layered structure, and evolving cyclic network structure [19, Figure 4]

Figure 4 provides a topological representation of the distributed memory system depicted in the right portion of Figure 2. It can be viewed from three perspectives:

1) The distributed memory is structurally organized in three layers. The top layer is controlled by words, consisting of simple one-dimensional array of symbols, logically constructed language, grammars that specify language use, etc. The middle layer resides on the behavioral eco-network for the individual to generate consciousness. In this layer, one acquires the meaning of behavior in the social ecology. The bottom layer creates unconsciously controlled behavioral eco-networks for the individual. This is a cyclic network starting from PMD, towards MMD via BMD, and returning to PMD. The results of activities of motor neurons, that reflect the activation of MMD, are perceived by sensory neurons to cause activation of PMD, which constitutes a closed network of PMD, BMD, and MMD.
2) Activation in the memory network spreads according to the Functional Flow Structure in the order of PMD, WMD, RMD, BMD, and MMD. An activated portion of RMD corresponds to consciousness. The flow can stay at the same layer for a while. For example, a certain word activates a set of words, and then, they activate another set of words. Consciousness formed at RMD at some moment shifts to another consciousness. No action can be associated for these System 2 activities.
3) The memories are cyclic and evolve in a cumulative and irreversible way, characterized by Evolving Cyclic Network Structure. The activated state of MMD is reflected on the activities of motor neurons, followed by updated input via sensory neurons to cause activation of PMD. PMD is shared by the three layers and serves as a common source of activation. Repetitive use of specific combinations of memory frames would strengthen the connections between them, and evolve.

It is important to note that memory serves as a mechanism to establish synchronization. A rough sketch concerning how synchronization between the external world and the internal world, i.e., successive PCM processes, could be done can

 ISBN: 978-1-61208-705-4

be drawn by combining MHP/RT and Distributed Memory System in Figure 2. When sensory neurons are activated by external stimuli, Perceptual Information Processing fires relevant portions of memory followed by firing of connected networks in the current memory structure, and the activated portions of memory are available for Conscious Information Processing (System 2) and Autonomous Automatic Behavior Control (System 1) via resonance reaction for some amount of time, then, finally the output of System 1 is input to Behavioral Action Processing to carry out behavior by sending signals to the associated motor neurons to act in the real world. The movement of one's body part will initiate the next cycle of PCM processes. What is going on in the external world and what the self is behaving are internalized through perception, and the perceived world goes through a PCM process by activating relevant portions of memory and utilizing it via resonance mechanism. In this way, memory processes play an important role to bridge the gap between the perceived external world and the internal world of PCM cycles to have them go coherently, i.e., keeping them synchronized.

III. Cognitive and Behavioral Principle in Multimodal Interactions

The cyclic connection is critical to understand the relationship between behavior and memory. The three features shown in Figure 4 enable pipelining the processes. However, it has to be scheduled in such a way that the PCM processes work smoothly as the external environments change at their pace. This section derives cognitive and behavioral principles in multimodal interactions that should organize the pipelining processes. They would provide a firm perspective to understand otherwise complex multimodal interaction processes.

A. Four Processing Modes: Conscious/Unconscious Processes Before/After an Event

Experience associated with a person's activities is characterized by a series of events, each of which is recognized by a person consciously. When one looks at the cognitive architecture MHP/RT from *a particular event* that occurred at the absolute time T in order to answer the question what it is doing for the event, MHP/RT's behavior looks as if it works in one of *four modes* [6][20] at one time before and after the event at T as shown by Figure 5.

Two of the four modes concern the processes carried out *before* the event:

- System 2 Before Mode: In the time range of $T-\beta \leq t < T-\beta'$, where $\beta' \sim 500$msec and β ranges a few seconds to hours, and even to months, MHP/RT uses memory, WMD and/or RMD, for *consciously* preparing for what would happen in the future.
- System 1 Before Mode: In the time range of $T-\beta' \leq t < T$, it *unconsciously* coordinate motor activities to the interacting environment. This mode uses PMD, BMD, and MMD.

The other two modes concern the processes carried out *after* the event:

- System 1 After Mode: In the time range of $T < t \leq T+\alpha'$, where $\alpha' \sim 500$msec, MHP/RT *unconsciously* tunes the connections between sensory inputs and motor outputs for better performance for the same event in the future. This mode updates the connections at the bottom layer of Figure 4.
- System 2 After Mode: In the time range of $T+\alpha' < t \leq T+\alpha$, it *consciously* recognizes what has happened, and then, modifies memory concerning the event, where α ranges a few seconds to minutes, and even to hours. This mode modifies the connections at the middle layer of Figure 4. Note that, since language (knowledge concerning words) is not directly related with multimodal interaction, the top layer of Figure 4 remains intact in the situations this paper deals with.

It is important to note, however, that an experience represented as a series of consciously identified events by a person has to be regarded as the results of unrecognized unconscious activities: metaphorically speaking, consciousness in System 2 is one of tips of icebergs that appear above the sea level, and the tips are interrelated with each other via the unseen relationships established below the sea level in System 1. A system of icebergs develops in the natural condition of seawater (temperature, tidal currents, etc.) and atmosphere, which may not be trivial for all people. Apparently, congruent configurations of the tips of two iceberg systems at a certain moment do not assure that they are entirely congruent. They may evolve differently as time goes by even if the surrounding environment is identical.

In summary, this subsection suggests that an understanding of the phenomena, that users are interacting with multimodal environment, could be obtained by regarding the phenomena as a series of conscious events, which could be further decomposed into four processing modes of MHP/RT concerning each event.

B. Representing Multimodal Interaction Event by Using the Four Processing Modes

Any event in multimodal interaction can be viewed from a user as an event that has happened in the time range of $[T-\beta,\ T+\alpha]$, and this should be an appropriate representation for the system event that occurs at T, accompanied with the portions of multi-dimensional memory frames that have participated in the processes in the time range of $[T-\beta,\ T+\alpha]$. It would be useful to consider a situation where a multimodal interface is expected to have the user integrate positive past experiences that are activated by provision of appropriate external cues at T from the system. The integrated memory formed in System 2 After Mode for the particular system event that happened at T would be activated in the future in System 2 Before Mode while preparing for the identical system events to happen repeatedly. The event appears recursively in the PCM processes, and is likely to extend the relevant time range longer, i..e., β and α would become larger. The larger β becomes, the farther the person can foresee. The larger α becomes, the wider the person can elaborate on the event. These should make the person smarter in living.

It is important to notice the fact that consciousness concerning the event comes to play in System 2 After Mode after the event at $T+\alpha'$ implies that consciousness lags at least by the amount of time, α', behind the real world. More specifically, the user is consciously blind during the period of $[T-\beta',\ T+\alpha']$ but would integrate consciously the blind period into conscious activities during the time

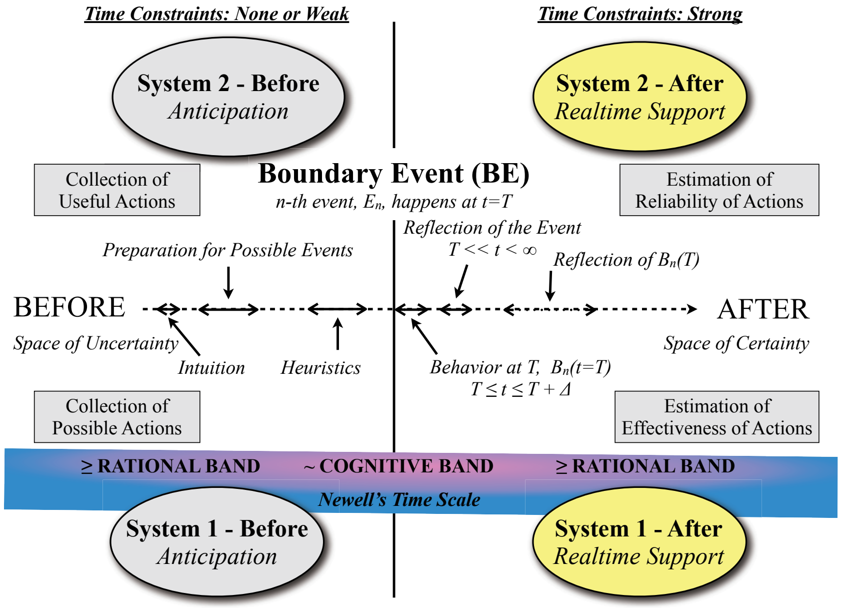

Figure 5. How the Four Processing Modes work (adapted from [20])

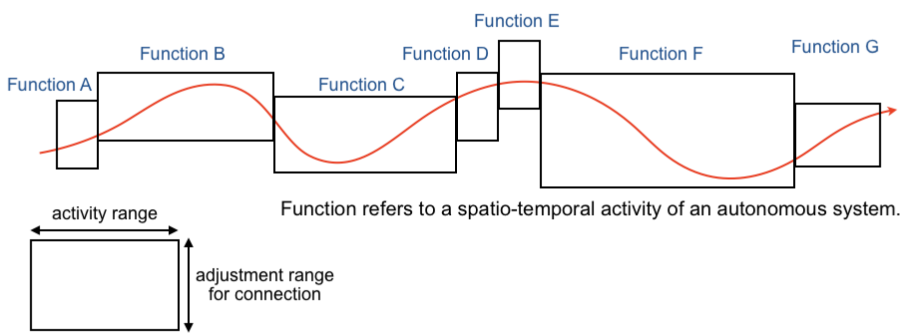

Figure 6. Successive functions are connected within the adjustable band in the spatio-time dimension.

range of $[T - \beta,\ T + \alpha]$. Therefore, consciously retrievable multimodal interaction experience has to be considered as memory structure concerning the event at T in the extended time range of $[T - \beta,\ T + \alpha]$ with consciously *inaccessible* but integrated memory region corresponding to the consciously blind time range of $[T - \beta',\ T + \alpha']$.

C. Weak Synchronization

Normally, the term "synchronization" refers to co-occurrence of two events on two distinct streams at the same time. In the case of multimodal interaction, the one side is a multimodal system and the other side is a user. It is said this way: a system and a user is synchronized if every system event at T_{sys} occurs as a user event at T_{user} with some amount of time allowance of Δ, $|T_{\text{user}} - T_{\text{sys}}| < \Delta$, where the actual values of Δ depend on the nature of interactions.

An example is as follows: a system sounds auditory cues, then, it shows text messages on a secondary display for 10 seconds with the expectation that its user hears the sound, comprehends it as the indication of something important to be shown on the secondary display for about 10 seconds, and moves his/her eyeballs to the target within, say, $2 \sim 3$ seconds. In this way, the system event and its corresponding user event are combined together to form a synchronized event at the overt information exchange level.

However, as depicted in Figure 5, a person's activity related with an event has to be considered from the four processing modes, which ranges relatively long time before and after the actual time the event happens. Therefore, "synchronization" has to be considered alternatively as the phenomena a person's activities during the time range of $[T - \beta, T + \alpha]$, which are linked with the specific recognizable system event at time T through a sequence of processes carried out in either of the four processing modes: all the processes have some link with the system event at T. When this is satisfied, the event is considered synchronized with a person's activities, which is called *weak synchronization* [21].

A smooth flow of the four processing modes can break when a person has to adjust his/her activity while s/he is in System 1 Before Mode in such a way that his/her movement goes in synchronous with the current environment. When this happens, the condition for weak synchronization is not satisfied but s/he has to make efforts to establish weak synchronization by adjusting his/her movement. When s/he reflects on this event in System 2 After Mode, he/she would have a feeling associated with anticipation-violated [21].

D. How MHP/RT Works under Weak Synchronization

In order for an O-PDP system, which is a higher level concept of MHP/RT, to engage in a particular system event at T during the time range of $[T-\beta, T+\alpha]$, the MHP/RT processes have to be chained (combined each other) to form a procedure, which include those to be performed during the respective time ranges of $[T-\beta, T-\beta')$, $[T-\beta', T)$, $(T, T+\alpha']$, $(T+\alpha', T+\alpha]$ for accomplishing appropriate functions in the ecological system composed of the multimodal system and the user. A mechanism is needed for establishing chains between functions that exist quasi-independently and discretely.

An O-PDP system is composed of autonomous elements in a band-structure, i.e., Biological, Cognitive, Rational, and Social Bands as suggested by Newell [16], and the processes carried out in the respective bands are combined each other to form a function. For example, in the Rational Band, a person could perform a series of inferences by logical deduction. Each inference is regarded as an element to obtain a logical result starting from an initial premise.

Autonomous elements are weakly synchronized with the external world, and the way how actually they work indirectly reflects the circularity of the existing environment – autopoiesis [12], and fluctuations inherent in the environment. This situation is schematically shown by Figure 6. Function C is connected with Function D using the region of the overlapping edge for maintaining continuity of the activities. Function C could be a series of conscious activities performed in the Rational Band to plan ahead a sequence of actions for controlling the car by consulting the contents of RMD (see Figure 4) followed by Function D, which could be an unconscious activity for tuning the planned activities for the particular road conditions by using the bottom layer of the memory structure, PMD, BMD and MMD, for which activations come from the middle layer following the Functional Flow Structure depicted in Figure 4. Function C is carried out in System 2 Before Mode in the time range of $[T-\beta, T-\beta')$ and Function D in System 1 Before Mode in the time range of $[T-\beta', T)$.

Note that elements in Biological, Cognitive, Rational, and Social Bands work autonomously. Conscious processes carried out in the upper bands run in parallel with unconscious processes performed in the lower band within the time range of $\leq$ 500msec. In Figure 6, Function C is a segment of an entire series of working of conscious elements that happens to establish connections with Function B, a segment from another conscious elements that was carried out before Function C, and with Function D, which is part of unconscious activities carried out by using the cycle of PMD, BMD, and MMD. The red curved line symbolically shows a trajectory of network firing in the O-PDP system, that is structured as Figure 4 by indicating the regions to connect to elements in different bands or those used for different processing modes. Various combinations of functional chains between parallel processes can occur, causing path proliferation of the network. This is possible because there exists *time relativity* among a variety of functions due to network circularity and fluctuation in processing in the behavioral ecology network of the O-PDP system. When the recall rates of specific paths become higher, proliferation along these paths are suppressed to centralize the activations on these paths thereafter.

An O-PDP system is created as a developed form of naturally formed energy circulation. Weak synchronization is the mechanism, or the principle, for a cognitive architecture, O-PDP system, to survive in the environment. Each element of the O-PDP system plays a certain role in achieving the overall goal of the whole O-PDP system. However, its role is not determined from the beginning but it exists only as the result of each element's own efforts to survive since it started its activity. The way of synchronization between elements is incidentally determined, and the synchronization itself is not deterministic but incomplete and flexible weak synchronization.

IV. Conclusion and Future Work

Whatever multimodal interaction environments users are in, what they would do for selecting next actions is simple: users just perform the dual-process of unconscious and conscious processes for every consciously recognizable event in certain time ranges before and after the event. Then, why is reasonable stability maintained in the uncertain procedure dependent on the autonomous reaction of the O-PDP system? In the situation of multimodal interaction, "stability" connotes the situations where the interactions between the system and the user continues smoothly without any breakdowns. The reasons are as follows:

- Perceptual input is used to generate a mirror image of the real world surrounding the self and shared in three bands, which are related with each other through the Functional Flow Structure and the Evolving Cyclic Network Structure (Figure 4).
- An O-PDP system is formed in the nonlinear hierarchically structured bands [16] to select next actions by applying four processes described in Section III-A as a means to survive in the ever-changing environment (Figure 5).
- Time constraints from the external environment is satisfied by fluctuations in the characteristic times of autonomous activities in respective bands through the mechanism of weak synchronization (Figure 6).

The first two items in the above list show the important characteristics derived from the architectural definition of O-PDP system, and the last item, weak synchronization, defines the principle for coordinating O-PDP in the ever-changing real environment. These three items jointly define the cognitive and behavioral principles necessary to understand users interacting with multimodal environment.

A. Future Work

The actual values of β and α, which define the time range of weak synchronization, can vary depending on the particular event at T, the contents of memory of the user, and the amount of time the user is allowed to allocate for the event. However, they can be estimated by simulating user behavior by MHP/RT. The purpose of simulation is to derive initial hypotheses concerning distinguishable users' behaviors caused by qualitatively different workings of MHP/RT, and the structure and contents of memory in the multimodal interaction environment. The hypotheses will be field-tested by having *selected users* representing the different segments in terms of characteristic behaviors in the particular multimodal interaction environment carry out the interactions. This approach, called *Cognitive-Chrono Ethnography (CCE)*, has been successfully applied to a variety of fields [4][5][22][23].

On the theoretical basis of the principles this paper described, some case studies are on-going to understand users' activities in multimodal interactions. Two of them are briefly described as future work that will come next to this work.

1) Designing Memorable Events: People live in the environment filled with artifacts, part of which is real and the rest is virtual. An initial theoretical simulations have been conducted to understand how the PCM processes along with the memory process result in memorable experiences [24]. Preliminary experiments were conducted to see how omnidirectional movies in virtual reality augmented with audio-guide made the experience memorable by timely *weak* synchronization and integration of multi-modal information. The contents of audio-guide for giving explanations to the visual contents to come after a few seconds have to be consciously processed by the usr in System 2 Before Mode. Some visual cues may be used to have the user unconsciously moves his/her eyeballs to the visual contents just having been given explanations, which is done in System 1 Before Mode. If this process happens while the memory is active, there is a good chance of strengthening the memory in System 2 After Mode, to cause the event memorable.

2) Designing Immersive Events: Immersive virtual environments are distinct from other types of multimedia learning environments. Initial theoretical considerations were reported that focused on the conditions necessary to produce "immersive experience" for the user [21]. Immersive feeling eliciting condition for an artificial environment to have the user feel immersive-ness is 1) it must be new to him/her, i.e., the range of memory activation is limited, 2) s/he is able to carry out actions with an anticipation activated by the artificial environment without any breakdown in performing motor-level actions System 2 Before Mode followed by no serious adjustment requited in System 1 Before Mode), 3) s/he is able to consciously recognize an event associated with the series of just-finished actions, and 4) s/he is able to reflect on the event to integrate it with the recognized feeling associated with the event (in System 2 After Mode). The study will continue in the context of developing a multimodal interface to help young pedestrians acquire necessary skills for safe navigation in dangerous traffic environments.

References

[1] J. L. Rubio-Tamayo, M. Gertrudix Barrio, and F. García García, "Environments and virtual reality: Systematic review and advances in communication, interaction and simulation," Multimodal Technologies and Interaction, vol. 1, no. 4, 2017.

[2] H. A. Simon, The Sciences of the Artificial, 3rd ed. Cambridge, MA: The MIT Press, 1996.

[3] M. Kitajima and M. Toyota, "Hierarchical Structure of Human Action Selections – An Update of Newell's Time Scale of Human Action," in Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures, vol. 41, 2014, pp. 321–325.

[4] ——, "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.

[5] M. Kitajima, Memory and Action Selection in Human-Machine Interaction. Wiley-ISTE, 2016.

[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] S. K. Card, T. P. Moran, and A. Newell, The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

[8] J. L. McClelland and D. E. Rumelhart, Parallel Distributed Processing: Explorations in the Microstructure of Cognition : Psychological and Biological Models. The MIT Press, 6 1986.

[9] D. Kahneman, "A perspective on judgment and choice," American Psychologist, vol. 58, no. 9, 2003, pp. 697–720.

[10] ——, Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux, 2011.

[11] D. Morris, The nature of happiness. London: Little Books Ltd., 2006.

[12] H. Maturana and F. Varela, Autopoiesis and Cognition (Boston Studies in the Philosophy and History of Science), softcover reprint of the original 1st ed. 1980 ed. D. Reidel Publishing Company, 8 1991.

[13] M. Kitajima and M. Toyota, "Two Minds and Emotion," in COGNITIVE 2015 : The Seventh International Conference on Advanced Cognitive Technologies and Applications, 2015, pp. 8–16.

[14] J. R. Anderson and C. Lebiere, The Atomic Components of Thought. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

[15] J. R. Anderson, How can the Human Mind Occur in the Physical Universe? New York, NY: Oxford University Press, 2007.

[16] A. Newell, Unified Theories of Cognition (The William James Lectures, 1987). Cambridge, MA: Harvard University Press, 1990, page 122, Fig. 3-3.

[17] 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, B. C. Love, K. McRae, and V. M. Sloutsky, Eds. Austin, TX: Cognitive Science Society, 2008, p. 2288.

[18] M. Kitajima and M. Toyota, "Distributed Memory System Architecture based on the Analyses of Human Brain Memory," in Proceedings of the 35th Annual Conference of the Cognitive Science Society, M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth, Eds. Austin, TX: Cognitive Science Society, 2013, p. 4002.

[19] ——, "Topological Considerations of Memory Structure," in Procedia Computer Science, BICA 2014. 5th Annual International Conference on Biologically Inspired Cognitive Architectures, vol. 41, 2014, pp. 45–50.

[20] ——, "Four Processing Modes of *in situ* Human Behavior," in Biologically Inspired Cognitive Architectures 2011 - Proceedings of the Second Annual Meeting of the BICA Society, A. V. Samsonovich and K. R. Jóhannsdóttir, Eds. Amsterdam, The Netherlands: IOS Press, 2011, pp. 194–199.

[21] J. Dinet and M. Kitajima, "Immersive interfaces for engagement and learning: Cognitive implications," in Proceedings of the 2015 Virtual Reality International Conference, ser. VRIC '18. New York, NY, USA: ACM, 2018, pp. 18/04:1–18/04:8. [Online]. Available: https://doi.org/10.1145/3234253.3234301

[22] M. Kitajima, H. Tahira, S. Takahashi, and T. Midorikawa, "Understanding tourist's *in situ* behavior: a Cognitive Chrono-Ethnography study of visitors to a hot spring resort," Journal of Quality Assurance in Hospitality and Tourism, vol. 12, 2012, pp. 247–270.

[23] M. Kitajima and M. Toyota, "Cognitive Chrono-Ethnography: A Methodology for Understanding Users for Designing Interactions Based on User Simulation with Cognitive Architectures," in Biologically Inspired Cognitive Architectures 2012 - Proceedings of the Third Annual Meeting of the BICA Society, 2012, pp. 193–198.

[24] M. Kitajima, S. Shimizu, and K. T. Nakahira, "Creating memorable experiences in virtual reality: Theory of its processes and preliminary eye-tracking study using omnidirectional movies with audio-guide," in 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), June 2017, pp. 1–8.

 ISBN: 978-1-61208-705-4