Basic Senses and Their Implications for Immersive Virtual Reality Design

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

Jérôme Dinet
Université de Lorraine
Nancy, France
Email: jerome.dinet@univ-lorraine.fr

Anissa Bystrzycki
Université de Lorraine
Nancy, France
Email: anissa.bystrzycki4@etu.univ-lorraine.fr

Anne-Marie Eby
Université de Lorraine
Nancy, France
Email: anne-marie.eby6@etu.univ-lorraine.fr

Marie Gérard
Université de Lorraine
Nancy, France
Email: marie.gerard1202@gmail.com

***Abstract*— Experiencing immersion in daily life, interacting with other agents, holds significant value. The object of the interaction may be an agent in the real world or one in the virtual world. Such experiences occur when sensory stimuli input from the five senses are integrated in the brain, and cognitive processing based on the results of that integration leads to the expression of selected behaviors, which are then accompanied by changes in the environment. A sense of immersion is created by the seamless integration of changes in the environment and in one's own state. In this study, we leverage the Model Human Processor with Realtime Constraints (MHP/RT), a cognitive architecture that can simulate human action selection and encompasses perceptual, cognitive, motor, and memory processes, and we argue that within the integration of multimodal sensory stimuli, *P-Resonance* – a resonance occurring between environmental information and perceptual memory, and *C-Resonance* – a subsequent resonance involving cognitive processes characterized by Two Minds and memory – are pivotal components contributing to the sense of immersion. The state of the external world is perceived as M-dimensional information by the five senses. This is passed to cognitive processing via P- and C-Resonances, converted into N-dimensional motor information in memory. Further, it is physically executed in the external world. We argue that the basic senses of rhythm, space, and number exist as mechanisms for these resonances to occur. It is shown that the spatial sense involving vision, not functioning before human birth, is integrated after birth with the rhythmic sense, which had been developing before birth. Furthermore, the number sense involving higher-order object recognition is further integrated with growth, which suggests that these basic senses are involved in the basis for the expression of human behavior. We contend that a virtual environment that does not deviate from the basic sensory-based mechanisms, which manifest interactions in the real world, is essential for the generation of immersive experiences.**

Keywords*— *Basic senses; Rhythmic sense; Spatial sense; Number sense; Resonance.

I. INTRODUCTION

The environment in which humans interact includes 1) *natural objects* that are direct objects of interaction, 2) *real artifacts* created by humans that exist as entities, 3) *virtual artifacts* that do not exist as entities but are perceived by humans through the illusion of their existence, and 4) real objects and artifacts that exist naturally and are indirectly involved in interaction as their background. Of these four categories, the third, "virtual artifacts," is rapidly permeating our lives. This change is attributed to the advancement of artificial intelligence algorithms, which now enable the creation of intricate illusions.

An interface exists between humans and the environment. When the interface is conscious, or visible, the human consciously performs actions toward the environment and consciously evaluates the response of the environment to those actions. In this case, the interface is visible to the human. Conversely, there are cases in which interaction proceeds without the human being aware of the interface. In this case, the interface is transparent or invisible, the human is immersed in the environment. Moreover, the human and the environment are seamlessly integrated.

When two interacting systems are seamlessly integrated, the processes in each system proceed in unison and in step with each other. The two systems are then synchronized. When both systems are linear, synchronization is easy because the progress of each process can be controlled based on highly accurate predictions. However, if either system is nonlinear, synchronization is challenging. Notably, human behavior can be captured by four nonlinearly connected hierarchies or bands, that is, biological, cognitive, rational, and social bands [1]; thus, humans are nonlinear behavior-generating systems. In the case of action selection, it can be captured by Two Minds [2][3], comprising a conscious system and an unconscious system. The conscious system is called System 2. This is a slow system that operates by feedback control, which executes inference using knowledge, whereas the unconscious system is called System 1 and is a fast system that operates by feed-forward control using intuition.

The behaviors that humans produce in their interactions with

 ISBN: 978-1-68558-154-1

the environment are the result of integrating the behaviors of System 1 and System 2, which have different characteristic times. Regardless of whether the system existing as an environment is a linear or nonlinear system, it is necessary to keep pace with human behavior to create a sense of immersion on the part of humans. This can be achieved by establishing a seamless relationship with the nonlinear human system.

When evaluating the alignment of the environment and humans states, it is not appropriate to simply consider that the actions of both systems are in sync with each other at time T. For an interaction event occurring at a certain time, humans consciously plan in advance and unconsciously coordinate and execute their actions just before the interaction. They also unconsciously make adjustments in the neural networks related to the generation of the action after the interaction. Further, they consciously reflect on the interaction when the outcome of the interaction is obtained. The conscious and unconscious learning after the interaction is used to plan actions when similar situations are encountered in the future [4]. From the human perspective, this event occurs within the time range of the Two Minds process operating before and after the event. Therefore, it is reasonable to view the synchronization between the environment and humans involved in the event generation within that time range. This type of synchronization with a time width is called *weak synchronization*. Further, it is proposed as a necessary condition for the generation of a sense of immersion. *Immersive feeling eliciting condition* for an artificial environment to have the user feel immersive-ness is defined as follows [5]:

IMMERSIVE FEELING ELICITING CONDITION

1) It must be new to them.
2) **With an anticipation activated by the artificial environment, they are able to perform actions without any breakdown in performing motor-level actions.**
3) They are able to consciously recognize an event associated with the series of just-finished actions.
4) They are able to reflect on the event to integrate it with the recognized feeling associated with the event.

The longer the state shown in Item 2 of the immersive feeling eliciting condition, the longer the immersive sensations can be felt without conscious intervention by System 2. Here, a series of actions is generated by feedforward control by System 1 as a series that is executed unconsciously without evaluation of the execution results. Let us assume that the first action is executed at time T and the last action in this series is executed at time T'. Subsequently, the interaction event generated by this action series is denoted by $E(T : T')$. The period $[T,\ T']$ when the action is executed unconsciously is the time when the action does not come to consciousness. For instance, the time when the action is interacting with the environment under feedforward control by System 1 without conscious intervention by System 2, and it can be regarded as the immersion time. The longer this time is, the longer the immersion is held.

The conditions for the establishment of weak synchronization that enables a continuous sense of immersion in the interaction with virtual artifacts depend on the perceptual, cognitive, and motor characteristics of the individual, including the content of memory, and the individual characteristics of the combination of these characteristics. This study examines how immersion is continuously maintained assuming general individual characteristics, based on Model Human Processor with Realtime Constraints (MHP/RT) [4][6]. MHP/RT is a comprehensive theory of action selection and memory, and provides the basis for building any model to understand human everyday behavior, including cognitive mechanism. Then, based on this, we discuss some individual characteristics that are unique but should be appropriately supported, and examine means to generate a sense of immersion.

This paper is organized as follows: Section II describes the outline of information uptake by perceptual processes from the external and internal environment, memory activation and execution of cognitive and motor processes through resonance in Section II-A. Section II-B introduces an important problem of binding and integrating multiple sensory information to produce behavior synchronized with spatiotemporal changes in the environment. Section II-C argues that the basic senses, that is, the rhythmic, spatial, and number senses, are critical for incorporating external information through resonance to make it available for the cognitive and motor processes to follow. Section III discusses the basic senses from the perspective of human development, including the prenatal period. Section IV concludes this paper by suggesting the points to be considered when designing virtual reality that smoothly connects to human activities.

II. INTERACTING WITH THE ENVIRONMENT USING RESONANCE

Satisfying the second item of Immersive Feeling Eliciting Condition (IFEC) is important for seamless interactions to continue to run as a feedforward process by System 1. The key for understanding the human–environment interaction process based on MHP/RT is that the communication between autonomous systems is achieved by a mechanism of *resonance* [7]. Both environmental systems and human systems are autonomous systems; human systems include perception, cognition, movement, and memory. This section describes how the continuous feedforward processing behind seamless interactions is supported by resonance mechanisms.

A. Interaction with the Environment Through Memory, Perception, Cognition, and Motor Processes Using Resonance

When interacting with the environment, humans respond to physical and chemical stimuli emitted from the external and internal environment by sensory nerves located at the interface with the environment and take in environmental information in the body. The brain acquires environmental information concerning the current activity of the self through the multiple

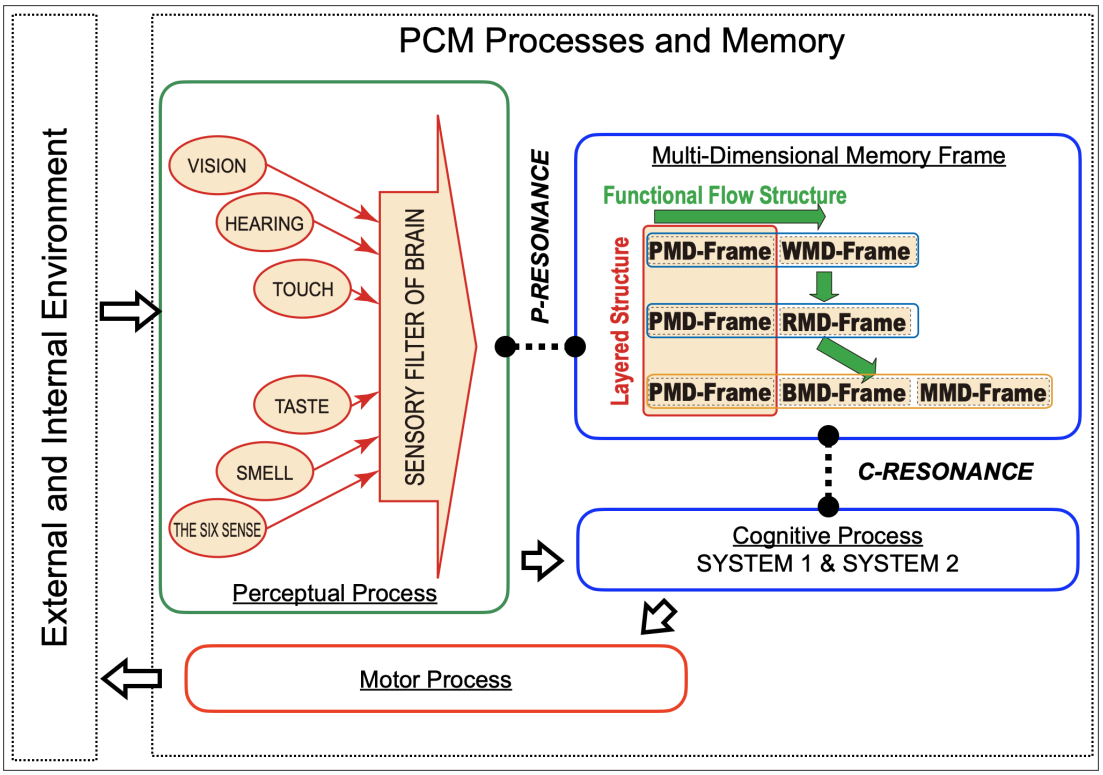

Figure 1. Information uptake by perceptual processes from the external and internal environment, memory activation and execution of cognitive and motor processes through resonance.

sensory organs. Further, it generates bodily movements that are suitable for the current environment. The stable and sustainable relationship between the environment and the self is established through continuous coordination between the activity of the self and the resultant changes in the environment, which should affect the self's next action.

Figure 1 shows the process, based on MHP/RT [6][8], by which environmental information, internal and external, 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 involves memory, which is modeled as Multi-Dimensional Memory Frame, and perceptual, cognitive (Two Minds), and motor processes. The memory structure, Multi-Dimensional Memory Frame, consists of Perceptual-, Behavior-, Motor-, Relation-, and Word-Multi-Dimensional Memory Frame. Perceptual-Multi-Dimensional Memory Frame overlaps with Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame, for spreading activation from Perceptual-Multi-Dimensional Memory Frame to Motor-Multi-Dimensional Memory Frame.

Perceptual information taken in from the environment through sensory organs resonates with information in the memory network structured as Multi-Dimensional Memory Frame, that is, P-Resonance. In Figure 1, this process is indicated by •—•. Resonance occurs first in the Perceptual-Multi-Dimensional Memory Frame within the Multi-Dimensional Memory Frame and activates the memory network. After that, the activity propagates to the memory networks that overlap the Perceptual-Multi-Dimensional Memory Frame, which are the Behavior-, Relation-, and Word-Multi-Dimensional Memory Frame, and finally to the Motor-Multi-Dimensional Memory Frame. In cognitive processing by Two Minds, conscious processing by System 2, which utilizes the Word- and Relation-Multi-Dimensional Memory Frame via C-Resonance, and unconscious processing by System 1, utilizing 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, which is the result of cognitive processing. The memories involved in the production of a behavior are updated to reflect the traces of its use process and influence the future behavior selection process.

B. Integration of Multiple Sensory Information to Produce Behavior Synchronized with Spatiotemporal Changes in the Environment

Information from the environment is taken in by multiple sensory organs. The sensory organs are located in different parts of the body with a certain extent, and the information collected is spatially distributed. Additionally, the information received by the sensory organs is time-series information. Therefore, the information taken in through the sensory organs is characterized as information that is spatially and temporally spread out. *How, then, do the structured Multi-Dimensional*

 ISBN: 978-1-68558-154-1

Memory Frame and perceptual information resonate? Or how can P-Resonance occur?

This question is known as the binding problem. It is a question of following form: *How does the brain integrate different aspects of perception into a coherent experience?* or *How does it perceive an object as a single entity while having distinguishable aspectual features such as shape, color, and texture?*

1) Dennett's Thoughts on the Binding Problem: Related to this question, Daniel Dennett, author of "Consciousness Revealed" [9], was asked by an interviewer, "You've talked about consciousness as something that we feel that's really more marvelous than it really is. Why is it?" To this, he responded as follows:

> *Yes, I think that consciousness plays tricks on us. Behind the eyes and in between the ears and the inner witnesses watching this wonderful show. But then when you do the, the physiology and you study perception, you realize it now in fact, you have a very limited take, you're only taking sips from them firehose of information that's coming in a little bit from vision a little bit from from hearing, and there's in fact this, this competition going on tug of war between different senses, between different interpretations of what you're seeing. All of this competition resolves itself in the fullness of time and pretty darn quick to produce the behavior that we're capable of and the reflection that we're capable of. It seems though, as if there's, it all comes together at some place.* ***And that's just an illusion. There's no place in the brain where it all comes together, for enjoyment and for and for witnessing by an inner witness.*** *So we have to take all that work that that inner witness was going to do, and we have to break it up in little bits and distributed around in the brain in the time and space that's available. And no one of those little bits is going to be conscious, and yet the sum of all of that work. Witnesses are going to do that with consciousness.*

Subsequently, he adds the following to the "illusion that everything flows together as one":

> *We have that illusion, and it is an illusion, because in fact, there's lots of things going on at once. They're not all that coherent.* ***And so our brains are very good at creating the sorts of simplifications that make that make its own job easier.*** *And so, yes, consciousness is an illusion of the brain for the brain, by the brain, if you like.*

2) Mechanistic Answers to the Binding Problem: Our answer to the binding problem is "sensory filter processing of multiple perceptual information and memory activation via *P-Resonance*." In the theorization of human perceptual, cognitive, and motor processes and memory by MHP/RT, the interface between the environment and the brain is performed by P-Resonance in perceptual information processing. Moreover, the interface between hierarchies in non-linearly connected hierarchical information processing within the brain is performed by C-Resonance. Resonance expresses the relation between two sides connected in a nonlinear relationship. Furthermore, it is a mechanism to transfer information that is ordered on one side to be treated as ordered information on the other side by mapping information. Notably, each information spreads with a specific manner on a common time axis [10].

C. Theoretical Derivation of Basic Senses

MHP/RT can be thought of as implementing the method described by Daniel Dennett in Section II-B1 on a system consisting of perceptual, cognitive, and motor processes and memory. Given the characteristics of memory described in Section II-B2, this section discusses perceptual memory, which is activated by P-Resonance. Further, it shows that *basic senses* exist to process perceptual information from sensory organs in an orderly manner. Basic senses include *rhythmic sense* related to time, to be described in II-C6a, *spatial sense* related to spatial perception, to be described in II-C6b, and *number sense* related to object segmentation in space, to be described in II-C6c.

1) Information that Perception Takes in: There are two types of information that perception takes in, which is represented as a discrete sequence of points spread over a time axis:

1) Stationary periodic sampling data; and
2) Differential information when it senses changes in the environment.

Moreover, there are two types of perception that involve different sources of information:

1) Perception for environmental information outside the body; and
2) Perception for monitoring the state of activity inside the body.

Therefore, the objects to be perceived can be broadly classified into four types depending on whether the data is stationary or non-stationary, and whether the source is inside or outside the body. There are various perceptual organs; each collecting its own perceptual information within its own perceptible bandwidth.

2) Memory Generation from Perceived Information: The perceived information concerns generation of memory for bodily behavior. Initially, memories are formed for different organs, and even within a single cell, in different forms. Nevertheless, human bodily behavior in general can be understood almost entirely through the workings of the central nervous system, which comprises the brain and spinal cord.

However, it is known from genetic analysis that the underlying structure of the nervous system is a chain of sensory, intervening, and motor nerves. An important function of the perception originated from sensory nerves is the identification of external objects to be transmitted to intervening nerves. The information from various perceptions is integrated to confirm the existence of the object and to store it in memory. Therefore,

 ISBN: 978-1-68558-154-1

most memories that reside in the cerebral cortex in brain are object memories.

3) $M \otimes N$ Mapping in Memory: The memory formed by the chain structure of sensory, intervening, and motor nerves stores the procedures of actions performed in response to events in the intervening stage, situated between perception and motor movement, by means of pairwise relations between perceptual components of the objects expressed as the M-dimensional information, that is, *perceptual objects*, and motor components of the objects expressed as the N-dimensional information, that is, *motor objects*. These procedures expressed in the intervening nerves can be viewed as chains in the Behavior-Multi-Dimensional Memory Frame. These are collectively referred to as "$M \otimes N$ mapping in memory" in MHP/RT [11].

4) Two Minds: System 1 and System 2: In the range of everyday behavior, processing is known to occur in the basal ganglia, that is, a group of neuronal nuclei that connect the cerebral cortex to the thalamus and brain stem, and below, where the $M \otimes N$ mapping in memory is coordinated by Two Minds. System 1 and System 2 work in parallel. The System 1 process is mainly responsible for physical activities, whereas the System 2 process works as an intervention for the purpose of coordinating activities. The role of System 2 is to intervene effectively in physical actions. This is accomplished by enhancing perceptual abilities through reconstruction of perceptual memory organized as Perceptual-Multi-Dimensional Memory Frame. For this purpose, System 2 consciously reflects on the results of physical actions conducted for accomplishing then-activated goals and reconstructs Perceptual-Multi-Dimensional Memory Frame according to the evaluation.

5) Representing Perceptual and Motor Objects in Relativized Time and Circulation Networks: The perceptual information taken in by sensory organs *does not contain absolute positional and temporal information*. The only shared feature among the various types of perceptual information is *simultaneity with other perceptual information* in the same time axis. Specifically, parallelism and cross-synchronization between each perceptual information processing are ensured. Additionally, it is possible to execute actions coherently as a procedure by integrating various types of perceptual information in a certain period of time on the time axis.

Human behavior forms a cyclic life ecology called circadian rhythm with a fluctuating bandwidth, which is caused by the stable day–night periodicity of the earth's rotation, the inclination of the axis of rotation with respect to the sun's orbital plane, and the slightly elliptical movement of the orbital circle. Human behavior contains recursive elements in which the results of one's actions are returned to oneself. Therefore, System 1 can act adaptively and flexibly in the next similar situations, whereas System 2 can obtain new effective action procedures by reflecting on the results of the action and organizing the perceptual information accordingly to reconstruct the existing memory network, that is, the Perceptual-Multi-Dimensional Memory Frame. In the future, System 2 intervenes in the sequence of actions generated by System 1 when it is needed to change its direction based on the reconstructed memory.

Human behavior can be viewed as an adaptive behavior for survival in the ever-changing and pseudo-cyclic environment. From birth to death, human beings act ceaselessly, which is a cyclic activity encompassing circadian rhythms. Hence, the neural network concerned with the execution of System 1 forms a circulatory network. Humans transform their behavior as their environment changes and as they themselves change with age. Accordingly, the initial circulatory network develops into a circulatory network with more complex connections and connection bandwidths by connecting the neural circuits involved in the execution of System 2 as procedures. Conscious procedure performed by System 2 happens in the operation and unit task systems in the cognitive band, and rational and social bands [1]. In MHP/RT, such procedures are represented in the Relation-Multi-Dimensional Memory Frame as consciously accessible relations of manipulable objects.

6) Memory Reuse: Rhythm-Based Reconstruction of Cognitive Objects: The earth is almost spherical. However, the range of human activity is far smaller than the global scale. Therefore, it is perceived as a three-dimensional Cartesian space rather than a coordinate system associated with a sphere.

Memories do not contain absolute temporal and three-dimensional positional information as described in Section II-C5. Meanwhile, System 1 and System 2 have to generate timely behavior appropriate for the time-position dependent situation. It is necessary to reconstruct the information from the memories by incorporating the time and positional information to make the time-position free memories available for System 1 and System 2 in the time position dependent ongoing situation. By doing this, System 1 and System 2 can reuse the time-position free memories at the point in time when it becomes necessary.

In Figure 1, the Multi-Dimensional Memory Frame and System 1 and System 2 are connected via C-Resonance, that is, the cognitive process incorporates the portion of memory through resonance that has been transformed into manipulable forms in the current situation by binding the time and position information in P-Resonance, which is called *Cognitive-Objects* in MHP/RT [12]. It is considered that, whereas perception of the external situation resonates with the Perceptual-Multi-Dimensional Memory Frame via P-Resonance, *the time-position free constructs* in the Perceptual-Multi-Dimensional Memory Frame are somehow modified to generate *the time-position dependent constructs*, that is, Cognitive-Objects, that are necessary to instantiate the real actions stored in the Motor-Multi-Dimensional Memory Frame via Relation- and Behavior-Multi-Dimensional Memory Frame. Then, how can it be possible? The key is the concept of *rhythm*, which characterizes the timing of the occurrence of an event. That is to say "binding positional and time information to the time-position free information by means of rhythm."

a) Rhythmic Sense: The changes brought about by actual human action are micro changes on cyclic activity. These

 ISBN: 978-1-68558-154-1

minute changes alter the relative situation between the actors themselves and the environment that involves others. However, from the three-dimensional understanding of human perception, these changes are perceived as continuous changes along the time axis. Conversely, if we look at the organs active in the human body, they have evolved and developed under circadian rhythms. Consequently, periodically active organs such as the heart have been formed to provide unique rhythms.

Turning to the environmental side, changes with various reproducible rhythms occur under the cyclic activity of the earth. Thus, to adapt to changes in the environment, a "rhythmic sense of basic perception" should be formed in the connection circuit of the circulatory network formed by various procedural memories. This is called rhythmic sense that enables flexible binding of memory and perceptual information on the time axis in P-Resonance, with hearing as the core and perception in general.

b) Spatial Sense through Rhythmic Sense: Bodily activity includes movement that involves changing the position of one's own body part in the three-dimensional space. Recognition of the current situation of the three-dimensional space is necessary for constructing executable bodily activities from the information stored in the Motor-Multi-Dimensional Memory Frame, that is free from absolute positions. The unique dimensions associated with movement are distance and time, which are required to make the move. The time is associated with body's internal rhythms, which define the scale for measuring distance. Thus, the information concerning distance between objects in the external environment is conceived through the rhythm-based scale, that can be called "spatial sense." Furthermore, the scale can change overtime because human behavior changes its orientation and range of circulation as it grows. Consequently, the cyclic trajectory thus formed evolves into a complicated web reflecting the range of variations of movement; humans expand their activity bandwidth. Inevitably, "spatial sense of basic perception", which serves as the basis for spatial cognition, should be formed in the connection circuit of the circulatory network formed by various procedural memories, which works in P-Resonance.

c) Number Sense: For humans to select appropriate actions in a timely manner in an ever-changing environment, information related to quantitative comparisons such as larger or smaller for size, more or fewer for the number of objects, farther or closer for distance, and longer or shorter for duration is indispensable. When this information is combined with the reward response that reflects the appropriateness of the choice, a basic sense of quantitative discrimination is formed. This is called number sense [13]. Perceptual information is represented in M-dimensional information in the Perceptual-Multi-Dimensional Memory Frame via P-Resonance with rhythmic and spatial senses. This process takes place in the midst of synchronization between the environment and human activity, which is weak synchronization, synchronization within the width of the time or activity bandwidth [5]. Subsequently, they are aggregated as cognitive objects by the number sense and made available to Two Minds via C-Resonance.

d) Summary: The rhythmic sense, spatial sense, and number sense are applied to generate Cognitive-Objects from the resonated portion of the Perceptual-Multi-Dimensional Memory Frame for further cognitive processing carried out by System 1 and System 2. The accuracy of adaptive behavior can be increased by repeating the behavior while developing the rhythmic, spatial, and number senses in the reuse of the Perceptual-Multi-Dimensional Memory Frame.

III. The Role of Resonance in the Human Development

Following birth, infants must immediately process and rapidly adapt to the array of unknown sensory experiences associated with their new ex-utero environment. However, and as Dall'Orso and colleagues (2020) [14] said, although it is known that unimodal stimuli induce activity in the corresponding primary sensory cortices of the newborn brain, it is unclear how multimodal stimuli are processed and integrated across modalities.

Despite the relatively immature state of their sensory modalities, newborns have perceptions that are acquired, and are triggered by, their contact with the environment. More recently, the study of the fetal origins of the sensory modes has revealed that in utero all the senses prepare to operate, except for the vision mode, which is only functional starting from the first minutes after birth [15][16]. This discrepancy between the maturation of the different senses leads to the question of how human newborns come to understand our multimodal and complex environment.

According to us, and as we wrote previously in this paper, the information taken from the sensory organs is characterized as information that is spatially and temporally spread out and we hypothesize that the Multi-Dimensional Memory Frame and perceptual information resonate (P-resonance) can be relevant to describe how multimodal stimuli are processes and integrated.

Just after the birth, the human newborn nervous system must also quickly adapt to process the temporal, spatial, and contextual features of complex inputs which simultaneously stimulate different sensory systems. This integration between different sensory modalities is thought to be established through associative learning during postnatal life as it allows simple and fast encoding of environmental contingencies [17].

If this integration between different sensory modalities has multiple roles in early development for biological and physiological aspects, for example, through anticipating and overcoming respiratory or thermal challenges during sleep, it plays also a key role for cognitive and social development such as behavioral and emotional-self-regulation [18], attention to facial and vocal expressions [19], linking different auditory–visual features to make inferences about specific objects [20][21], and facilitating development of language and vocabulary growth [22].

Dall'Orso and colleagues (2020) [14] have demonstrated that the early human brain is already capable of processing

 ISBN: 978-1-68558-154-1

external and simultaneous multisensory information within the two distinct primary sensory cortices. We further found that this multisensory stimulus presentation can directly influence cortical activity in a crossmodal manner, even in the absence of its archetypal substrate. They also found that the process of encoding information during associative learning in the neonatal period engages wider cortical association regions across relevant neural networks. Recent Magnetoencephalography (MEG)/Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) studies have suggested that while posterior parietal cortices integrate signals weighted by their sensory reliabilities irrespective of task context, anterior parietal cortices encode spatial estimates depending on their reliability and task relevance [23][24][25]. In anterior parietal cortices, spatial estimates rely more on the location of the signals of the sensory modality that needs to be reported. While these behavioral and neuroimaging findings demonstrate that observers' perceptual goals influence how the brain combines sensory signals to support perceptual inference, the underlying computational and neural mechanisms remain controversial and we argue that MHP/RT can provide a relevant theoretical framework to explain these mechanisms.

And finally, their results confirmed prior results obtained by other authors who identified the primary sensorimotor cortices as the encoders of our associative learning task, supporting the idea that they are not only involved in basic sensory perception but also higher level processes such as multisensory integration and experience specific memory encoding [26]. Moreover, some recent studies investigating multimodal integration for newborns with perceptual impairments, such as visual impairments [27][28][29] and for children with Autism Spectrum Disorder, or ASD [30], provide very similar results.

IV. Conclusion and Future Work

This study argued in Section II-B that binding position and time to the Multi-Dimensional Memory Frame is essential to generate situation-adaptive behavior, and derived in Section II-C basic senses for solving the binding problem. The memory represented by the Multi-Dimensional Memory Frame, which has no position and time data, represents relationships between objects. Therefore, the contents of the memory can be reused by performing a topological transformation to match the current time and spatial scales. Furthermore, the accuracy of current perceptual information can be enhanced based on past memories and associated time and space values. This is because the Perceptual-Multi-Dimensional Memory Frame is simply a memory of object relationships; the Behavior-Multi-Dimensional Memory Frame remembers object relationships associated with the time of the entity's actions in the procedures of physical behavior; the Relation-Multi-Dimensional Memory Frame stores object relationships associated with procedures and temporal concepts recognized by System 2.

Section III discussed the basic senses from a developmental perspective, tracing it back to the prenatal stage. It also clarified how we are able to process multimodal information now that we are living surrounded by various artifacts, which implements immersive virtual reality. Based on the argument this paper provided, establishing P-Resonance between the external environment and the Perceptual-Multi-Dimensional Memory Frame with the focus of the basic senses. These include the rhythmic sense, spatial sense, and number sense, should be critical for smooth integration of the environment and human being in action. As long as the source of stimuli to human sensory organs, whether virtual or real, is not smoothly processed by the basic senses, it will not be taken into the organism and processed in relation to it. If those stimuli are Immersive Virtual Reality generated by applying Artificial Intelligence, they must be generated to achieve some goal, but if they are not compatible with the basic senses, that goal cannot be achieved.

In everyday life human beings experience a continuous stream of information that they perceive through sight, sound, smell, taste, and touch. Even though this experience is mostly multisensory, that is, they receive information from multiple senses simultaneously, psychological research has primarily focused on studying our senses in isolation. Multisensory processing refers to the interaction of signals arriving nearly simultaneously from different sensory modalities. This implies that information from one modality can influence information processing in another modality. Information from different sensory modalities can also be combined into a single multisensory event, a process that is referred to as multisensory integration [31]. By this way, IVR is used to create virtual worlds that are as immersive as possible in order to make users feel as if they are "really there", immersion referring to the objective capacity of the technology to deliver sensorial stimulations and movement tracking, for example, head, hands, comparable to their physiological manifestations in the physical world [32]. It is assumed that the more immersed someone feels while exposed to a virtual world, then the more memorable the IVR experience is specially for children [33]. Even if some studies show that too much visual stimulations, for example, by a larger field of view or more visual details, can be responsible for negative effects and discomfort due to increased eye strain on the part of the user, that is, cybersickness, the quality of immersion results from the combination of various factors, such as field of view (the extent of the visible world), head tracking, visual fidelity (the realism and details of visual information) and multisensory information, which is thought to be a central element in immersion.

At the moment of sampling continuous events in space–time, which is the moment the memory is formed, absolute time and coordinate information is lost and associations between events are made with features that can be mapped to other events. Therefore, when the events are drawn out at the time of motion, spatio-temporal information between events is necessary, so it is interpolated to generate it. P-Resonance occurs between the external stimulus and the Perceptual-Multi-Dimensional Memory Frame by rhythmic and spatial senses, and M-dimensional perceptual representations are generated. Subsequently, under the condition of weak-synchronization,

 ISBN: 978-1-68558-154-1

cognitive objects are generated by utilizing the number sense. C-Resonance occurs between the cognitive object and the Two Minds, and finally, mapped to the N-dimensional representation of the Motor-Multi-Dimensional Memory Frame. The N-dimensional motion representation is interpolated in space–time and converted into motor-enabled information to generate motion via motor nerves. The body plan, that is, skeleton, supports the whole behind the scenes as the base of the interpolation program as the default value. For memories created in a real environment to be effective in a VR environment, which would be IVR, VR design should be based on these characteristics.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 20H04290. The author would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

[1] A. Newell, Unified Theories of Cognition (The William James Lectures, 1987). Cambridge, MA: Harvard University Press, 1990.

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

[3] D. Kahneman, Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux, 2011.

[4] M. Kitajima and M. Toyota, "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.

[5] 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

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

[7] J. Dinet and M. Kitajima, "The Concept of Resonance: From Physics to Cognitive Psychology," in COGNITIVE 2020 : The Twelfth International Conference on Advanced Cognitive Technologies and Applications, 2020, pp. 62–67.

[8] 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.

[9] D. C. Dennett, Consciousness Explained (English Edition). Little, Brown and Company, 2 2017.

[10] M. Kitajima et al., "Language and Image in Behavioral Ecology," in COGNITIVE 2022 : The Fourteenth International Conference on Advanced Cognitive Technologies and Applications, 2022, pp. 1–10.

[11] 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.

[12] 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.

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

[14] S. Dall'Orso et al., "Cortical Processing of Multimodal Sensory Learning in Human Neonates," Cerebral Cortex, vol. 31, no. 3, 11 2020, pp. 1827–1836. [Online]. Available: https://doi.org/10.1093/cercor/bhaa340[retrieved:January,2023]

[15] A. Streri, M. D. d. De Hevia, V. Izard, and A. Coubart, "What do we know about neonatal cognition?" Behavioral Sciences, vol. 3, no. 1, 2013, pp. 154–169. [Online]. Available: https://www.mdpi.com/2076-328X/3/1/154[retrieved:January,2023]

[16] A. Streri and M. D. de Hevia, "How do human newborns come to understand the multimodal environment?" Psychonomic Bulletin & Review, vol. 30, no. 4, Aug. 2023, pp. 1171–1186. [Online]. Available: https://doi.org/10.3758/s13423-023-02260-y[retrieved:January,2023]

[17] K. Connolly, "Multisensory perception as an associative learning process," Frontiers in Psychology, vol. 5, 2014. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01095[retrieved:January,2023]

[18] L. P. Lipsitt, "Learning and Emotion in Infants," Pediatrics, vol. 102, no. Supplement_E1, 11 1998, pp. 1262–1267. [Online]. Available: https://doi.org/10.1542/peds.102.SE1.1262[retrieved:January,2023]

[19] A. S. Walker-Andrews, "Infants' perception of expressive behaviors: Differentiation of multimodal information." Psychological Bulletin, vol. 121, no. 3, 1997, pp. 437–456, place: US Publisher: American Psychological Association.

[20] B. A. Morrongiello, K. D. Fenwick, and G. Chance, "Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events," Infant Behavior and Development, vol. 21, no. 4, 1998, pp. 543–553. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0163638398900285[retrieved:January,2023]

[21] Y. Tanaka, Y. Kanakogi, M. Kawasaki, and M. Myowa, "The integration of audio − tactile information is modulated by multimodal social interaction with physical contact in infancy," Developmental Cognitive Neuroscience, vol. 30, 2018, pp. 31–40. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1878929316301232[retrieved:January,2023]

[22] L. Smith and C. Yu, "Infants rapidly learn word-referent mappings via cross-situational statistics," Cognition, vol. 106, no. 3, 2008, pp. 1558–1568. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027707001795[retrieved:January,2023]

[23] T. Rohe and U. Noppeney, "Cortical hierarchies perform bayesian causal inference in multisensory perception," PLoS biology, vol. 13, no. 2, 2015, p. e1002073.

[24] T. Rohe and U. Noppeney, "Distinct computational principles govern multisensory integration in primary sensory and association cortices," Current Biology, vol. 26, no. 4, 2016, pp. 509–514.

[25] M. Aller and U. Noppeney, "To integrate or not to integrate: Temporal dynamics of hierarchical bayesian causal inference," PLoS biology, vol. 17, no. 4, 2019, p. e3000210.

[26] D. B. Headley and N. M. Weinberger, "Relational Associative Learning Induces Cross-Modal Plasticity in Early Visual Cortex," Cerebral Cortex, vol. 25, no. 5, 11 2013, pp. 1306–1318. [Online]. Available: https://doi.org/10.1093/cercor/bht325[retrieved:January,2023]

[27] N. J. Dale et al., "Home-based early intervention in infants and young children with visual impairment using the developmental journal: longitudinal cohort study," Developmental Medicine & Child Neurology, vol. 61, no. 6, 2019, pp. 697–709. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/dmcn.14081[retrieved:January,2023]

[28] F. Morelli et al., "A multidimensional, multisensory and comprehensive rehabilitation intervention to improve spatial functioning in the visually impaired child: A community case study," Frontiers in Neuroscience, vol. 14, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2020.00768[retrieved:January,2023]

[29] A. Gatecel and A. Valentin-Lefranc, Le Grand Livre des pratiques psychomotrices. Fondements, domaines d'application, formation et recherche, ser. Hors collection. Paris: Dunod, 2019. [Online]. Available: https://www.cairn.info/le-grand-livre-des-pratiques-psychomotrices--9782100798940.htm[retrieved:January,2023]

[30] N. M. Salowitz, "Multimodal sensory integration for perception and action in high functioning children with autism spectrum disorder," Ph.D. dissertation, Marquette University, 2015.

[31] M. Quak, R. E. London, and D. Talsma, "A multisensory perspective of working memory," Frontiers in human neuroscience, vol. 9, 2015, p. 197.

[32] L. B. Cadet and H. Chainay, "Memory of virtual experiences: Role of immersion, emotion and sense of presence," International Journal of Human-Computer Studies, vol. 144, 2020, p. 102506.

[33] L. B. Cadet, E. Reynaud, and H. Chainay, "Memory for a virtual reality experience in children and adults according to image quality, emotion, and sense of presence," Virtual Reality, vol. 26, no. 1, 2022, pp. 55–75.

 ISBN: 978-1-68558-154-1