Basic Senses and Their Implications for Immersive Virtual Reality Design

Muneo Kitajima [1] (Presenter)   Makoto Toyota [2]   Jérôme Dinet [3]
Anissa Bystrzycki [3]   Anne-Marie Eby [3]   Marie Gérard [3]

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

[2] T-Method, JAPAN

[3] Université de Lorraine, FRANCE

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

Project HP: http://oberon.nagaokaut.ac.jp/ktjm/organic-self-consistent-field-theory/index.html

HP: http://oberon.nagaokaut.ac.jp/ktjm/index.html

email: mailto:mkitajima@kjs.nagaokaut.ac.jp

How the Sense of Immersion Created?

Experiencing immersion in daily life while 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.

This Study is About ···

We leverage the Model Human Processor with Realtime Constraints [1, 2], 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.

- 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.
- The behaviors that humans produce in their interactions with the environment are the result of integrating the behaviors of System 1 and 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.

IMMERSIVE FEELING ELICITING CONDITION [3]

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.

- This study examines how immersion is continuously maintained assuming general individual characteristics, based on Model Human Processor with Realtime Constraints (MHP/RT)* [4, 1].
- 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.

* 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.

- 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.
- ▶ The information received by the sensory organs is time-series information.
- ⇒ 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 Memory Frame and perceptual information resonate? Or how can P-Resonance occur?

Binding Problem

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?*

Related to this question, Daniel Dennett, author of "Consciousness Revealed" [5], 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.*

- 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, resonance plays an important role.
  ▶ The interface between the environment and the brain is performed by *P-Resonance* in perceptual information processing.
  ▶ 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.
  ▶ 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.
  ▶ Each information spreads with a specific manner on a *common* time axis [6].

■ **The first component (left) comprises cyclic PCM processes. They execute a series of events in synchronous with changes in the external environment.** The parallel distributed processing [7] for realizing these PCM processes is implemented as hierarchically organized bands introduced by Newell [8, Figure 3-3]. These bands are characterized by characteristic operation times, which are defined by associating relative times with individual PCM processes. Events occur by connecting what happens in a band to what happens in its adjacent band *non-linearly*. A mechanism is required to connect the events; MHP/RT suggests that this connection is provided by *the resonance mechanism* via the MDMFs, shown by •—• in the figure.

■ **The second component (right) is the autonomous memory system consisting of five MDMFs, which are Perception, Motion, Behavior, Relation, and Word MDMFs.** The MDMFs store information associated with the corresponding autonomous processes defined in the PCM processes. The MDMFs are subservient to the PCM processes because they do not exist unless the PCM processes do.
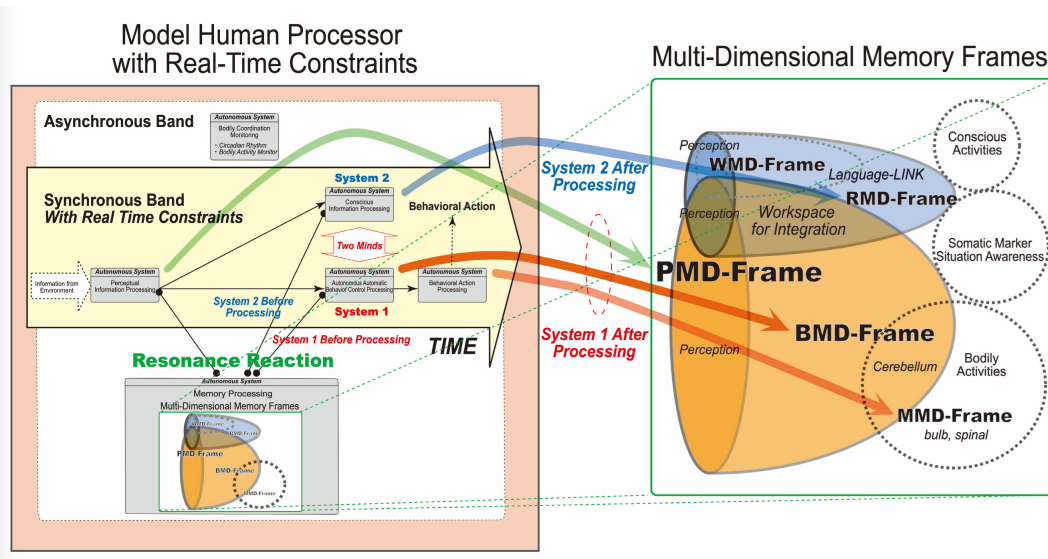
Four Operation Modes of MHP/RT

Synchronous Modes

Mode 1: System 1 driven mode

A single set of perceptual stimuli initiate feed-forward processes at the B-(iological) and C-(ognitive) bands to act with occasional feed-back from an upper band, i.e., C-, R-(ational), or S-(ocial) bands.

Mode 2: System 2 driven mode

A single set of perceptual stimuli initiate a feed-back process at the C-band, and upon completion of the conscious action selection, the unconscious automatic feedforward process is activated at the B- and C-bands for action.

Asynchronous Modes

**Mode 3: In-phase autonomous activity mode**

A set of perceptual stimuli initiate feedforward processes at the B- and C-bands with one and another intertwined occasional feedback processes from an upper band, i.e., C-, R-, or S-bands.

Mode 4: Heterophasic autonomous activity mode

Multiple threads of perceptual stimuli initiate respective feedforward processes at the B- and C-bands, some with no feedback and others with feedback from the upper bands, i.e., C-, R-, or S-bands.

Four Processing Modes of MHP/RT

The experience associated with an individual's activity is characterized by a series of events that are consciously recognized serially. Let $E(T_N)$ denote the event that occurred at time T_N. The experience is then defined as a series of events along the timeline, " $\cdots \rightarrow E(T_{N-1}) \rightarrow E(T_N) \rightarrow E(T_{N+1}) \rightarrow \cdots$ ". Considering the way System 1 and System 2 are involved in individual events, four processing modes can be defined:
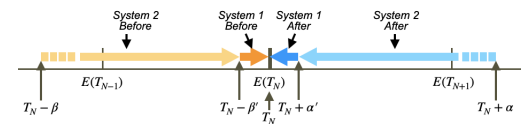
- **System-2-Before-Event-Mode**: In the time range of $T - \beta \leq t < T - \beta'$, MHP/RT plans for future events to occur. There is enough time to think carefully.
- **System-1-Before-Event-Mode**: In $T - \beta' \leq t < T$, the action selections smoothly generate the immediate event.
- **System-1-After-Event-Mode**: In $T < t \leq T + \alpha'$, to perform better for the same event in the future, the connection between the incoming perceptual information and the output motor content is adjusted unconsciously.
- **System-2-After-Event-Mode**: In $T + \alpha' < t \leq T + \alpha$, the event is reflected upon. The results are stored and used in the next System-2-Before-Event-Mode before a similar event occurs.

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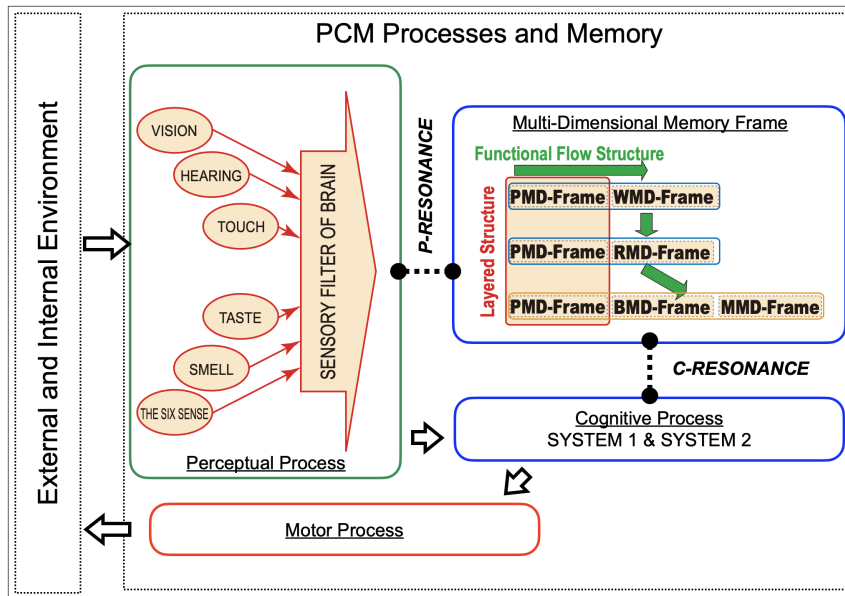
Figure 1: Information uptake by perceptual processes from the external and internal environment, memory activation and execution of cognitive and motor processes through resonance.

MHP/RT can be thought of as implementing the method described by Daniel Dennett on a system consisting of perceptual, cognitive, and motor processes and memory. Given the characteristics of memory, this slide shows 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.

[1] Information that Perception Takes in

Two types of information:
1. Stationary periodic sampling data; and
2. Differential information when it senses changes in the environment.

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

[2] Memory Generation

- Initially, memories are formed for different organs in different forms.
- Nevertheless, human bodily behavior can be understood through the workings of the central nervous system.
- An important function of the perception originated from sensory nerves is the identification of external objects to be transmitted to intervening nerves.

[3] $M \otimes N$ Mapping in Memory

- The memory stores the procedures in the intervening stage, situated between perception and motor movement.
  - ▶ Perceptual components of the objects expressed as the M-dimensional information.
  - ▶ Motor components of the objects expressed as the N-dimensional information.
- These procedures are stored as chains in the Behavior-Multi-dimensional Memory Frame.

[4] Two Minds: System 1 and System 2

- $M \otimes N$ mapping is coordinated by Two Minds.
- System 1 process is mainly responsible for physical activities.
- System 2 process works as an intervention for the purpose of coordinating activities.
  - ▶ This is accomplished by enhancing perceptual abilities through reconstruction of perceptual memory organized as Perceptual-Multi-dimensional Memory Frame.

[5] Representing Perceptual and Motor Objects in Relativized Time and Circulation Networks

- The perceptual information taken in by sensory organs *not containing absolute positional and temporal information*.
- ► The only shared feature is *simultaneity with other perceptual information* in the same time axis.
- Human behavior forms a cyclic life ecology with a fluctuating bandwidth; it contains recursive elements in which the results of one's actions are returned to oneself.
- ► System 1 can act adaptively and flexibly in the next similar situations;
- ► 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, i.e., the Perceptual-MDMF.
- ⇒ 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.

[6] Memory Reuse: Rhythm-Based Reconstruction of Cognitive Objects

- *Memories do not contain absolute temporal and 3-D positional information.* Meanwhile, System 1 and 2 have to generate timely behavior appropriate for the time-position dependent situation.
- ► *Reconstruction of the information from the memories is needed* by incorporating the time and positional information to make the time-position free memories available for System 1 and 2 in the time-position dependent situation.
- Perception resonates with the Perceptual-MDMF via P-Resonance, *the time-position free constructs* are modified to generate *the time-position dependent constructs*, Cognitive-Objects, necessary to instantiate the real actions stored in the Motor-MDMF.
- ► *How can it be possible?* The key is the concept of *rhythm*, which characterizes the timing of the occurrence of an event.
- ⇒ ***"Binding positional and time information to the time-position free information by means of rhythm."***

Rhythmic Sense of Basic Perception

- ■ Periodically active organs have been formed to provide unique rhythms.
- ▶ Environmental changes with reproducible rhythms occur under the cyclic activity of the earth.
- ⇒ To adapt to environmental changes , a "rhythmic sense" should be formed for flexible binding of memory and perceptual information in *P-Resonance*.

Spatial Sense through Rhythmic Sense

- ■ Bodily activity includes movement, i.e., changing the position of one's own body part in the 3-D space.
- ▶ Recognition of the current situation is necessary for constructing executable bodily activities from the information in the Motor-MDMF that is free from absolute positions.
- ■ The unique dimensions associated with movement are distance and time, required to make the move.
- ▶ The time is associated with body's internal rhythms, defining the scale for measuring distance.
- ⇒ The information concerning distance between objects in the external environment is conceived through the rhythm-based scale, that can be called "spatial sense."
- ⇒⇒ "Spatial sense of basic perception", as the basis for spatial cognition, should be formed in the circuit of the circulatory network formed by various procedural memories, which works in *P-Resonance*.

Number Sense

- ■ For humans to select appropriate actions in a timely manner in an ever-changing environment, information related to quantitative comparisons (e.g., larger or smaller for size) 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 [9].
- ■ Perceptual information is represented in M-dimensional information in the Perceptual-MDMF via P-Resonance with rhythmic and spatial senses.
- ▶ The perceptual representations are aggregated as cognitive objects by the number sense and made available to Two Minds via C-Resonance.

Newborn's Perception

- Following birth, infants must immediately process and rapidly adapt to the array of unknown sensory experiences associated with their new ex-utero environment.
  - ▶ It is unclear how multimodal stimuli are processed and integrated across modalities [10].
  - ▶ 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 [11, 12].

How do human newborns come to understand our multimodal and complex environment?

P-Resonance for Integration of Stimuli

- The information taken from the sensory organs is characterized as information spatially and temporally spread out.
  - ▶ *P-resonance* between the perceptual information and Multi-dimensional Memory Frame can be relevant to describe how multimodal stimuli are processed and integrated.
  - ▶ This integration is thought to be established through associative learning during postnatal life as it allows simple and fast encoding of environmental contingencies [13].

Effects of Integration on Development

- If this integration between different sensory modalities has multiple roles in early development for biological and physiological aspects, it plays also a key role for cognitive and social development such as;
  - – Behavioral and emotional self-regulation [14],
  - – Attention to facial and vocal expressions [15],
  - – Linking different auditory–visual features to make inferences about specific objects [16, 17],
  - – Facilitating development of language and vocabulary growth [18].

Conclusion

- This study argued that binding position and time to the MDMF is essential to generate situation-adaptive behavior; derived basic senses, i.e., rhythmic, spatial, and number senses, for solving the binding problem.

- The memory represented by the MDMF, which has no position and time data, represents relationships between objects.
- ⇒ *The contents of memory can be reused by performing a topological transformation to match the current time and spatial scales.*

- Establishing *P-Resonance* between the external environment and the Perceptual-MDMF with the focus of the basic senses should be critical for smooth integration of the environment and human being in action.
- This study discussed the basic senses from a developmental perspective, tracing it back to the prenatal stage and clarified how we are able to process multimodal information, living surrounded by various artifacts, which implements immersive VR.

Implications for Immersive VR Design I

- 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.
- 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 sensory stimulation and movement tracking, for example, head, hands, comparable to their physiological manifestations in the physical world [19].

Implications for Immersive VR Design II

- 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, *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 – 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.***

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

[2] Muneo Kitajima and Makoto Toyota. **Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT).** *Biologically Inspired Cognitive Architectures*, 5:82–93, 2013.

[3] Jérôme Dinet and Muneo Kitajima. **Immersive interfaces for engagement and learning: Cognitive implications.** In *Proceedings of the 2015 Virtual Reality International Conference*, VRIC '18, pages 18/04:1–18/04:8, New York, NY, USA, 2018. ACM.

[4] Muneo Kitajima and Makoto Toyota. **Four Processing Modes of *in situ* Human Behavior.** In Alexei V. Samsonovich and Kamilla R. Jóhannsdóttir, editors, *Biologically Inspired Cognitive Architectures 2011 - Proceedings of the Second Annual Meeting of the BICA Society*, pages 194–199, Amsterdam, The Netherlands, 2011. IOS Press.

[5] Daniel C. Dennett. ***Consciousness Explained (English Edition).*** Little, Brown and Company, 2 2017.

[6] Muneo Kitajima, Makoto Toyota, Jérôme Dinet, Clelie Amiot, Capucine Bauchet, and Hanna Verdel. **Language and Image in Behavioral Ecology.** In *COGNITIVE 2022 : The Fourteenth International Conference on Advanced Cognitive Technologies and Applications*, pages 1–10, 2022.

[7] James L. McClelland and David E. Rumelhart. ***Parallel Distributed Processing: Explorations in the Microstructure of Cognition : Psychological and Biological Models.*** The MIT Press, 6 1986.

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

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

[10] S Dall'Orso, W P Fifer, P D Balsam, J Brandon, C O'Keefe, T Poppe, K Vecchiato, A D Edwards, E Burdet, and T Arichi. **Cortical Processing of Multimodal Sensory Learning in Human Neonates.** *Cerebral Cortex*, 31(3):1827–1836, 11 2020.

[11] Arlette Streri, Maria Dolores de De Hevia, Véronique Izard, and Aurélie Coubart. **What do we know about neonatal cognition?** *Behavioral Sciences*, 3(1):154–169, 2013.

[12] Arlette Streri and Maria Dolores de Hevia. **How do human newborns come to understand the multimodal environment?** *Psychonomic Bulletin & Review*, 30(4):1171–1186, August 2023.

[13] Kevin Connolly. **Multisensory perception as an associative learning process.** *Frontiers in Psychology*, 5, 2014.

[14] Lewis P. Lipsitt. **Learning and Emotion in Infants.** *Pediatrics*, 102(Supplement_E1):1262–1267, 11 1998.

[15] Arlene S. Walker-Andrews. **Infants' perception of expressive behaviors: Differentiation of multimodal information.** *Psychological Bulletin*, 121(3):437–456, 1997. Place: US Publisher: American Psychological Association.

[16] Barbara A. Morrongiello, Kimberley D. Fenwick, and Graham Chance. **Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events.** *Infant Behavior and Development*, 21(4):543–553, 1998.

[17] Yukari Tanaka, Yasuhiro Kanakogi, Masahiro Kawasaki, and Masako Myowa. **The integration of audio⍰tactile information is modulated by multimodal social interaction with physical contact in infancy.** *Developmental Cognitive Neuroscience*, 30:31–40, 2018.

[18] Linda Smith and Chen Yu. **Infants rapidly learn word-referent mappings via cross-situational statistics.** *Cognition*, 106(3):1558–1568, 2008.

[19] Lénaïc B Cadet and Hanna Chainay. **Memory of virtual experiences: Role of immersion, emotion and sense of presence.** *International Journal of Human-Computer Studies*, 144:102506, 2020.