Model-Based Analysis of the Differences in Sensory Perception between Real and Virtual Space : Toward “Adaptive Virtual Reality”

Taichi Nakagawa
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: s223308@stn.nagaokaut.ac.jp

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

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

***Abstract*— Although the implementation of “Adaptive Virtual Reality” is becoming feasible, understanding the main effects of its realization on users based on cognitive models is essential. Here, as the first step, we first describe a model of the flow of information obtained by actual human perception through avatars in virtual reality (VR) and the resulting human reactions, and confirm the validity of the user models proposed so far. We also consider the degree of immersion predicted due to the integration of multimodal information. The cognitive processes of VR experiences are largely categorized into “perception and recognition of information (attention, memory, and decision making)” and “perception-based physical actions and interactions with VR objects”. Based on this, we describe a cognitive model of VR experiences. In addition, as examples of the discrepancies in sensory perception experienced in real/VR spaces, we briefly describe the phenomena that occur in communication. We describe the cognitive models for these phenomena and qualitatively consider the degree to which sensory information obtained from the real/VR space affects the degree of chunks activation. The intensity of human sense is expressed as a logarithm according to Weber-Fechner’s Law, suggesting that human senses can distinguish differences even with weak sensory information. We argue that the “slightly different from the real world” sense felt in VR content is caused by such slight differences in sensory information. Overall, we advance the cognitive understanding of the immersive experience particularly in the VR space, and qualitatively describe the possibility of designing highly immersive VR content which are adapted to each individual.**

Keywords*— *sensory perception; cognitive model; virtual reality; experience.

I. Introduction

With the growth in Virtual Reality (VR) goggles and low cost of equipment for shooting omnidirectional video, VR content has attracted substantial attention. In addition to games, a wide range of VR contents have been developed, including omnidirectional video playback, education, sightseeing, property previews, and shopping. VR systems that enable these contents to be viewed are also growing rapidly. For example, the following innovations have emerged in content design. VR systems using Head-Mounted Displays (HMDs) sold to general consumers cover the user’s field of vision; thus, the user cannot see their own body. Therefore, VR systems using HMDs typically display a virtual body drawn from the user’s first-person perspective. A mechanism for realizing the user’s first-person perspective is the implementation of avatars. The effects of avatars have been described by researchers. Steed et al. [1] suggested that the use of avatars that follow the user’s movements can reduce the cognitive load of certain tasks in the VR space. People around the world have been using VR social networking services, such as VRChat, where users enjoy interacting with other users using avatars that they have selected and edited to their liking. Theis shows that avatars are a means of self-expression in VR communication.

There are many research approaches to VR contents and systems, including research from the perspective of Human Computer Interaction (HCI), research on the differences in sensory perception between the real world and VR, and research on “Adaptive VR” that incorporates individual adaptability into VR contents.

Among the studies from the perspective of HCI, Mousavi et al. [2] integrate Emotion Recognition (ER) and VR to provide an immersive and flexible environment in VR. This integration can advance HCI by allowing the Virtual Environment (VE) to adapt to the user’s emotional state.

Research on the difference in sensory perception between the real world and VR can be broadly divided into two perspectives: research from the perspective of illusions, and the other is from a purely cognitive perspective, including the cognitive load of the Working Memory (WM). Studies from the perspective of illusions have existed for a long time, including many on real-world phenomena. The most famous examples include the illusion phenomenon Rubber Hand Illusion (RHI) reported by Matthew, Jonathan [3], and others, the possession illusion and Proteus effect proposed by Yee and Bailenson [4].

As an extension of the RHI, Slater et al. [5][6] indicated that it could be produeced for virtual hands on a screen. Sanchez-Vives et al. [7] indicated that a visuo-haptic synchronization stimulus can induce a possession illusion for a virtual hand and suggested that this illusion can be induced without using a tactile stimulus.

Next, among studies about Proteus effects via avatar include, Yee and Bailenson showed that the use of avatars with different levels of attractiveness and height in appearance changes the way people communicate with others. Similarly, Oyanagi et al. [8] found that the use of a dragon avatar in a VR space can reduce the fear of heights. In a study by Tacikowski et al. [9], participants were exposed to images of the opposite sex’s body through an HMD. Participants indicated that subjective and implicit aspects of their gender identity, and stereotypical images of the opposite sex changed when they felt a sense of possession of the opposite sex's body.

Next, among studies on the relationship between VR and WM, Chiossi et al. [10] considered the influence of WM load, which leads to over- and under-stimulation, in the design of VR space. The authors designed an adaptive system to support the WM task execution based on electroencephalography (EEG) correlations between external and internal attention.

In parallel, a concept called by "Adaptive VR" has been discussed in recent years. Baker and Fairclough [11] described it as follows: Adaptive VR monitors human behavior, psychophysiology, and neurophysiology to create a real-time model of the user. This quantification is used to infer the emotional state of individual users and induce adaptive changes within the VE during runtime. Therefore, the authors argued that the efficacy of the emotional experience can be increased by modeling individual differences in the way users interact within a particular VE as a system.

To realize a seriese of practical studies, we need to follow the VR technological trends. Currently, many common VR technologies are in practical use. The low cost of HMDs, such as the Meta Quest2, has made it possible for consumers to easily experience VR content. There are two types of VR content: those in which the user does not move much and does move substantially in the VR space. Examples of the former include watching video content and browsing the web. In this case, the user's movements are mainly button operations and cursor movements using a controller, and the user rarely moves in both the real and VR spaces.

Examples of the latter include VR games and VR Social Networking Service (SNS) such as VR Chat [12]. VR games include those in which the user's actual body movements, such as swinging a sword or boxing, are synchronized with the movements of the avatar in the VR space, and those in which the user can move around in the VR space by operating a controller. Avatars are usually used in contents that allow users to move around in the VR space. Avatars are 3D objects that serve as the user's body in the VR space. Indeed, the use of avatars improves the realism of the VR experience and decreases the cognitive load in the VR space.

Given this background, the implementation of adaptive VR is becoming feasible. However, the main effects of its implementation on users should be understood based on a cognitive model. Studies have mainly focused on bottom-up content design with an awareness of adaptive VR. However, it is difficult for empirical developments to provide effects that create new phenomena. Hence, not only a bottom-up but also a top-down approach is necessary. As a stepping stone to this goal, we do the following in this study. We describe a model of the flow of information obtained by actual human perception through avatars in VR and the resulting human reactions, and confirm the validity of the user models proposed so far. The degree of immersion predicted because of the integration of multisensory information is also discussed. Understanding the role of multisensory information can enable us to design VR contents for individual users and how we can control sensory perception.

The remainder of this article is organized as follows. We describe the sense-perception cognitive model on VR in Section II. Next, we present an example of the difference between real-world and VR. Based on these, we finally explore the perception in the real and virtual worlds.

II. Description of the Cognitive Model for Sensory in VR

In general, physical information in the VR space is represented as follows. Objects in the VR space (VR objects) are represented by computer graphics, and their behavior is based on a program previously written to interact with the environment and other objects. The sound in the VR space is provided by artificially preparing audio data that is predicted in advance to be uttered in the space, and is played continuously in a background music-like manner, or by using a sound engine controlled by the user. Specifically, in the latter case, it can be attached to a VR object and played when certain conditions are met. Comprised of these elements, all human activities and virtual experiences in the VR space are performed by using the avatar as one's own body The avatar's movement is performed by tracking the user's real-world body movements. Tracking methods include three-point tracking, which consists of an HMD and two hand controllers, and full-body tracking, which uses motion capture and a tracking suit.

Consequently, the human experience in the VR space differs slightly from perception and cognition in the real world, and can be said to be the result of the interaction between avatar and VR objects, as well as the perception of the accompanying environment such as sound linked to these objects. Considering this, the model of human perception, cognition, and behavior in the VR space should be described with an awareness of the various interactions in the VR space with those in the real world.

- Perception of information
- Cognition of information
  - Attention
  - Memory
  - Decision
- Body motion based on perception
  - Human body motion
  - Interaction with VR objects

A. Perception of information

In general VR experiences using current HMDs, visual, auditory, and somatosensory information are used as perceptual information. The VR experience begins when the user puts on the HMD and views the images displayed on the lenses; by moving their head while wearing the HMD, the user can perceive the virtual space in the same way as they perceive the real world. Auditory information is output from the HMD's built-in or external speakers, and audio is played in response to the behavior of VR objects. The somatosensory information is

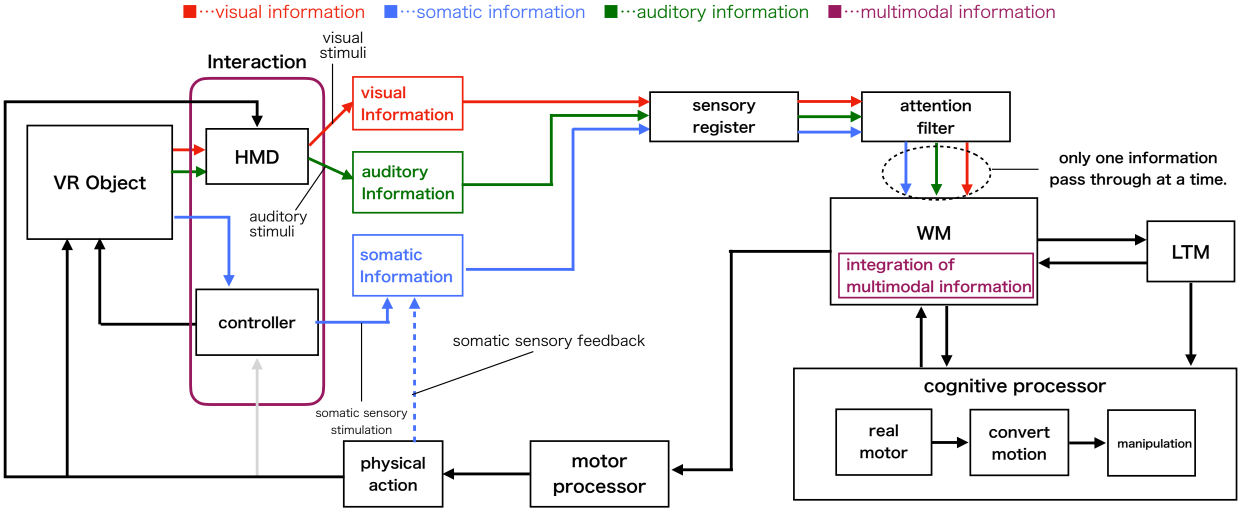

Figure 1. A Cognitive Model of VR Experience with HMD.

used to make operations in the VR space clearer by vibrating the controllers in both hands to generate tactile feedback when operating the User Interface (UI) in the VR space or selecting VR objects.

B. *Cognition of information*

1) Attention: Perceptual information moves to the sensory register, and then only the information to which the user's attention is directed passes through the selective filter and into the WM. Here, each sensory information does not completely enter the WM at the same time, but one piece of information passes through per processing.

2) Memory: If the sensory information obtained in the VR space is similar to that obtained in the real world, the user perceives the VR space as if it were a real space. In addition, based on the information in the Long-Term Memory (LTM), the user anticipates and expects the response of objects in the VR space to his or her actions, and engagement is generated.

3) Decision: Based on the perceptual information, the next action is determined. Here, when actions on a VR object are performed via a controller, the actions in the real world are converted into the corresponding controller operations.

C. *Body movement based on perception*

1) Human body movement: The operator (actual body) moves, and the avatar in the VR space moves in response to the movement. There are two methods for incorporating human motion into VR: (1) Image sensing by the camera attached to the HMD's basic UI operations (clicking and screen scrolling) and grasping VR objects (realized by holding something with a hand gesture) is possible. The high degree of synchronization between the actual hand and the avatar's hand motion is an advantage of this method. Conversely, precise manipulation, movements large enough to cause both hands to move out of the camera's field of view, and very fast hand movements are weaknesses. (2) Yaw, pitch, roll + relative position by controller. The accurate tracking of position, posture, and motion information by sensors is possible, and the sense of actual body motion is directly reflected during the operation, resulting in a high sense of immersion. However, if the reflection of body motion by the HMD is not synchronized with the actual body motion, it may cause a sense of discomfort and reduce the immersiveness of the VR experience.

2) Interaction with VR objects: VR objects not only appear to be three-dimensional, but can also be actually manipulated. Examples include playing a musical instrument or a push-button switch. Here, the immersiveness of the VR experience can be enhanced by providing not only a visual 3D effect, but also contextual information that one's actions affect the VR object.

D. *Integration of information*

Figure 2 shows the timeline of perceptual information in the WM when the perceptual information moves from the attention selector to the WM and activates information in the LTM from within the WM in Figure 1. Here, Information N refers to the information obtained from sensory organ N (e.g., vision). This information arrives in the WM at time t_N and exists for τ_N seconds. N is the number of perceived information. The long-term memory activated by these sensory information in the WM is denoted as h_M. In the case of Figure 2, there are two long-term memories activated by each of $N = 1, 2, 3$, and M is the number of i, j, k, l, p, and q in the WM. The time at which h_M arrives at the WM from the LTM is t_M, and the residence time is τ_M. Here, consider the contrast with perceptual information processing in the real space. For example, suppose that we now experience an event E_{rw} in real space. Suppose that H_i, h_j, h_k, h_l, and h_p of the LTM information are activated and processed simultaneously in the WM. Suppose that when an event E_{vw} close to E_{rw} is experienced in VR space, the information h_i, h_j, h_k, h_l, h_p in long-term memory is activated as a result of obtaining Information 1, 2, and 3, and processed simultaneously in the WM. Since the information processed in the WM is similar

ISBN: 978-1-68558-154-1

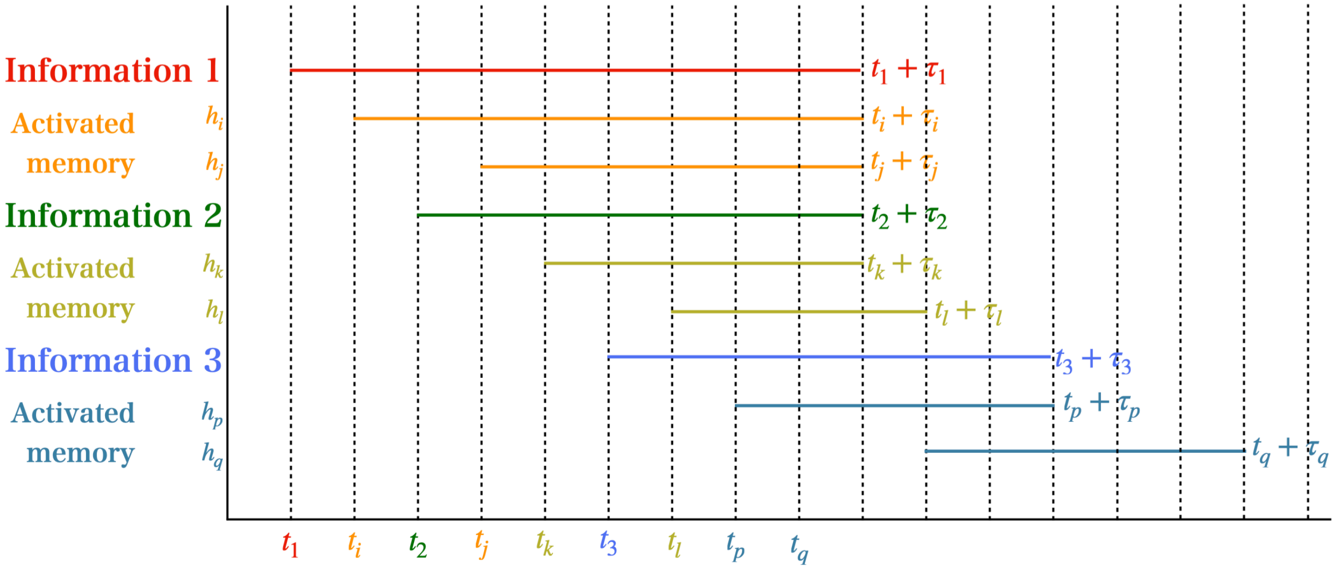

Figure 2. Staying timeline of sensory information stored in working memory and information invoked from long-term memory.

to that processed by the event E_{rw} in the real space, the VR experience is perceived as real and a sense of immersion is generated.

III. Examples of discrepancies between the real and VR worlds

A. Example 1: Playing Japanese Taiko

As an example, consider a situation in which the user plays *taiko* drums in a VR space. When the user sees a virtual *taiko* drum at time t_1, only visual information about the drum, I_1, exists in the user's WM. At time t_i, the sighting of the virtual *taiko* activates the *taiko* information h_i in the LTM, which becomes part of the information in the WM. Based on the user's experience in the real world, the user picks up the virtual stick and strikes the *taiko*, producing sound from the virtual *taiko*. At time t_2, the audio information reaches the user. Furthermore, the controllers of both hands vibrate, and somatosensory information reaches the user at time t_3. However, if the user knows from past experience that they feel air vibrations in their whole body when they hit the *taiko* drum, the information in the LTM does not match the information in the WM. This may cause a sense of discomfort and reduce the immersiveness of the VR experience.

B. Example 2: Communication within a VR space

Consider communication using avatars in a VR space. First, visual information, such as facial expressions and gestures of another avatar, exists in the user's WM at time t_1. Then, the voice of the other avatar reaches the user, and the voice information exists in the user's WM at time t_2. At this time, if the timing of the visual and audio information in the WM is off, such as if the other person's voice is heard from in front of the user even though the avatar of the conversation partner is behind the user, or if either of the two sensory information is unclear, the communication may feel uncomfortable or the immersiveness of the VR experience may be reduced.

IV. Perception in the Real/Virtual World

Based on Figures 1 and 2, we consider the perception of a phenomenon in the real (R) or virtual (V) space as follows. The chunk C_j stored in the LTM is constructed from the information group $I_i^{env}(t)$ obtained from sensory organ $i(1 \leq i \leq 5)$ in the past. Here, i refers to the five sensory organs possessed by a person. Each $I_i^{env}(t)$ passes through the attention filter $F_i^{env}(t)$ via the sensory register. And at time t, only the information obtained from a specific sensory organ passes through. C_j contains the information obtained from each sensory organ as a set $I(t)$ and is denoted as $C_j(I(t))$. Here, $I(t)$ is represented as follows:

$$I(t) = \{\boldsymbol{I} \; | I_i^{env}(t) F_i^{env}(t), \; 1 \leq i \leq 5\}.$$

The information that has passed through the attention filter is stored in the WM for a specific time, and a set of information $I(t)$ is sent to the LTM at the same time or with a time lag. In the LTM, $C_j(I(t))$ is matched with $C_j(I(t))$ based on the information in $I(t)$, and the closest or matching $C_j(I(t))$ is used as knowledge. The used knowledge is overwritten in the LTM through the WM in the form that the information in $I(t)$ is enhanced. Here, we target three sensory organs – visual, auditory, and somatic. We consider how the information flows through these three types of sensory organs in turn.

Suppose that at a certain time, a specific amount of information $I_i^{env}(t)(i = 1, \; 2, \; 3)$ is received from the external environment. $I_i^{env}(t)$ correspond to Information N in Figure

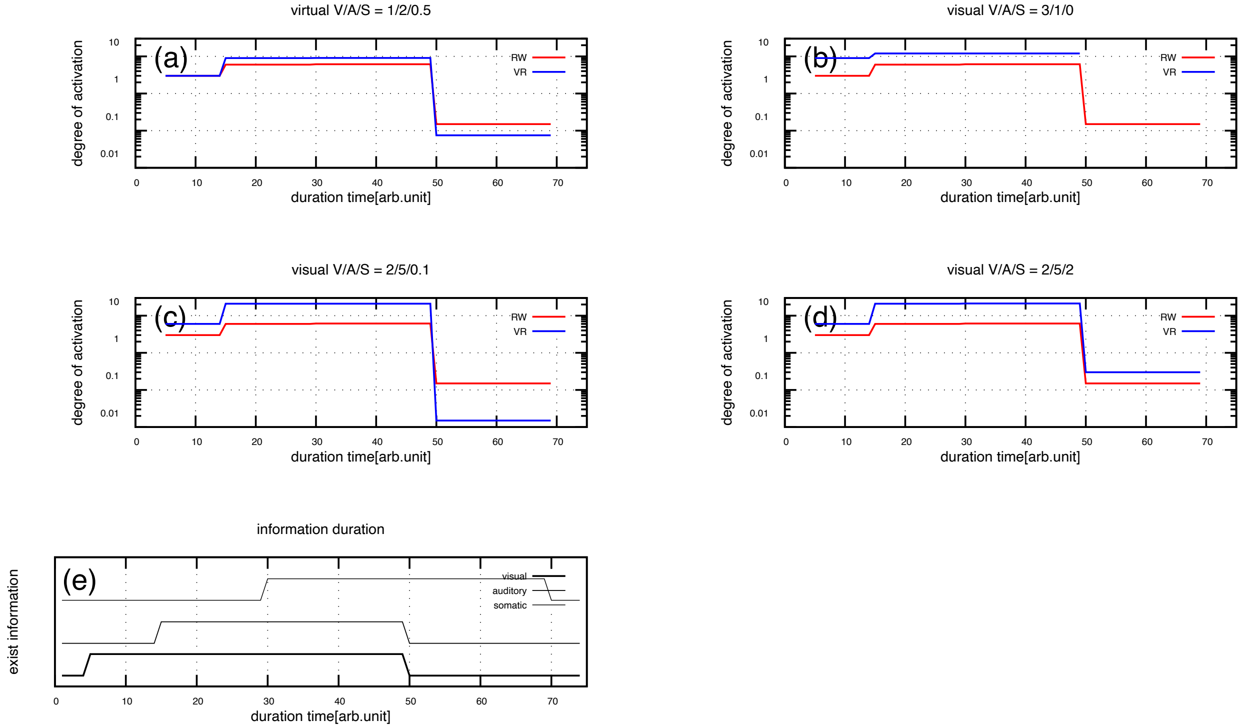

Figure 3. The trends of estimated $I^{syn}(t)$ which are changed three perceptual information(visual, auditory, somatic) amplified in Virtual Reality space.

2. Information N simultaneously activates several chunks. Although the degree of chunk activation varies, $I_i^{env}(t)F_i^{env}(t)$ is integrated into a single piece of information and sent to the LTM. In this case, the integration operator G can be used in various ways. The integrated information $I^{syn}(t)$ can be expressed as follows.

$$I^{syn}(t) = G(\ i,\ j,\ I_i^{env}(t)F_i^{env}(t),\ C_j(I(t))\ )$$

For the sake of simplicity, we simply add the amount of information and the degree of chunk activation as follows.

$$G^{env}(t) = \sum_i^n \sum_j^m I_i^{env}(t)F_i^{env}(t)C_j(I(t)) \qquad (1)$$

Figure 3 shows the trend of $I^{syn}(t)$ when the degrees to which visual, auditory, and somatic information are emphasized in VR are varied. The solid red line in the figure shows $I^{syn}(t)$ when visual, auditory, and somatic information are received in the real world. Here, we set $j = 1, 2$. Both visual and auditory information equal 1 for one, and 2 for the other. The somatic information is set to 0.5 on one side and 0.3 on the other. The solid blue lines indicate the degree to which the same information is distorted in VR.

Figure 3(e) shows the duration of information obtained from each sense. In contrast, Figures 3(a)∼(d) shows the degree of integrated information activation calculated by Equation(1). Figure 3(a) shows the case where auditory is multiplied by a factor of 2 and somatic by a factor of 0.5. For $t < 50$, the VR space is slightly more chunk activated, but the characteristics are almost same. However, at $t \geq 50$, when only somatic information is perceived, the chunk activation in the VR space is lower. In Figure 3(b), the visual information is markedly increased, while the somatic information is not reproduced in the VR space. For $t < 50$, the activation of chunk in the VR space is markedly increased, but at $t \geq 50$, the somatic information is lost; Hence, there is no chunk activation in the VR space. In Figure 3(c), the somatic information is lowered to 0.1 and the information is emphasized in the form of visual<auditory. In particular, at $t \geq 50$, the somatic information is still present, but its effect is much smaller. Figure 3(d) is the case where the somatic information is also doubled. Compared with Figure 3(b) and (c), chunk activation remains high at $t \geq 50$.

The intensity of human sensation is expressed as a logarithm according to Weber-Fechner's Law. Therefore, as shown in Figure 3, even if the difference in sensory information is very slight, it suggests that the human senses can distinguish this difference. The sense of "slightly different from the real world" felt in VR content is thought to be caused by such slight differences in sensory information. The sensory information obtained in real space is not necessarily large, as shown in the example in Section III. However, it is easy to understand that these small differences lead to a sense of discomfort, which in turn indicates a decrease in immersive perception.

In the present case, we only dealt with a very simple integration of information. To advance our understanding of human sensory perception and use knowledge in VR spaces, scholars should develop a new approach that uses operators, such as Adaptive Control of Thought—Rational (ACT-R) [13] and Model Human Processor with Realtime Constraints (MHP/RT) [13] which incorporate Two Minds, to integrate information in a cognitive architecture [14][15][16].

V. Conclusion and Future Work

To realize adaptive VR, we need to design deeper immersion resulting from human interaction with real/VR spaces. As a first step, this study describes a sensory-cognitive model for VR spaces. The model is based on the integration of multimodal information, and the relationship between the three types of sensory information (visual, auditory, and somatic) and chunk activation. To understand the actual phenomena based on the described model, we consider and analyze the example of communication in the VR space with *taiko*, and refer to what kind of discomfort is likely to occur and its underlying causes. Connecting the two issues, multimodal information and chunk activatin, we undertake qualitatively research and explain the phenomenon that can occur when one or more types of information (visual, auditory, or somatic) is overemphasized or surpressed in a VR space. Expressing human sensory intensity as a logarithm according to Weber-Fechner's Law, we suggest that human senses can distinguish differences in sensory information, even if the differences are very slight. Considering these points, we are able to deepen our understanding of how the VR space realizes the immersive effect with impressive each other. Moreover, we are able to design "adaptive" immersive contents. In the future, it is necessary to investigate in experiments whether the degree of immersion felt by users changes when they experience VR content by changing the degree of emphasis of each sensory information. The metrics used to judge the degree of similarity between the real and virtual worlds can be defined as the overlap between the information held in the WM and the information in the LTM that has been activated up to that point in time. As the activation of information in the LTM is considered to be reflected in biological information, future experiments could be conducted using eye gaze and skin resistance measurements and subjective evaluation by means of questionnaires. Hysteresis can be considered based on the impact of inputs from the environment on the memory of the time series.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 20H04290, 22K12284, 23K11334, and National University Management Reform Promotion Project. The authors would like to thank Editage (www.editage.com) for the English language editing.

References

[1] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in 2016 IEEE Virtual Reality (VR), 2016, pp. 67–76.

[2] S. M. H. Mousavi et al., "Emotion recognition in adaptive virtual reality settings: Challenges and opportunities," CEUR Workshop Proceedings, vol. 3517, jan 2023, pp. 1–20. [Online]. Available: https://sites.google.com/view/wamwb/

[3] B. Matthew and C. Jonathan, "Rubber hands 'feel' touch that eyes see," Nature, vol. 391, no. 6669, 02 1998, pp. 756–756. [Online]. Available: https://cir.nii.ac.jp/crid/1363388843275776640

[4] N. Yee and J. Bailenson, "The proteus effect: The effect of transformed self - representation on behavior," Human Communication Research, vol. 33, 07 2007, pp. 271–290.

[5] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. Sanchez-Vives, "Towards a digital body: the virtual arm illusion," Frontiers in Human Neuroscience, vol. 2, 2008. [Online]. Available: https://www.frontiersin.org/articles/10.3389/neuro.09.006.2008

[6] M. Slater, D. Pérez Marcos, H. Ehrsson, and M. Sanchez-Vives, "Inducing illusory ownership of a virtual body," Frontiers in Neuroscience, vol. 3, 2009. [Online]. Available: https://www.frontiersin.org/articles/10.3389/neuro.01.029.2009

[7] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater, "Virtual hand illusion induced by visuomotor correlations," PLOS ONE, vol. 5, no. 4, 04 2010, pp. 1–6. [Online]. Available: https://doi.org/10.1371/journal.pone.0010381

[8] A. Oyanagi, T. Narumi, and R. Ohmura, "An avatar that is used daily in the social vr contents enhances the sense of embodiment," Transactions of the Virtual Reality Society of Japan, vol. 25, no. 1, 2020, pp. 50–59.

[9] P. Tacikowski, J. Fust, and H. H. Ehrsson, "Fluidity of gender identity induced by illusory body-sex change," Scientific Reports, vol. 10, no. 1, 2020, p. 14385. [Online]. Available: https://doi.org/10.1038/s41598-020-71467-z

[10] F. Chiossi, C. Ou, C. Gerhardt, F. Putze, and S. Mayer, "Designing and evaluating an adaptive virtual reality system using eeg frequencies to balance internal and external attention states," 2023.

[11] C. Baker and S. H. Fairclough, "Chapter 9 - adaptive virtual reality," in Current Research in Neuroadaptive Technology, S. H. Fairclough and T. O. Zander, Eds. Academic Press, 2022, pp. 159–176. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128214138000142

[12] VRChat Inc., "Vrchat," [Online]. Available from:
https://hello.vrchat.com/, accessed, 2024.3.14.

[13] F. E. Ritter, F. Tehranchi, and J. D. Oury, "ACT-R: A cognitive architecture for modeling cognition," WIREs Cognitive Science, vol. 10, no. 3, 2019, p. e1488. [Online]. Available: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcs.1488

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

[15] 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.

[16] 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.