Visual Behavior Based on Information Foraging Theory Toward Designing of Auditory Information

Yuta Kurihara[1], Motoki Shino[2], Katsuko T. Nakahira[3], [a] and Muneo Kitajima[3] [b]

[1]*Department of Human & Engineered Environmental Studies, The University of Tokyo, Kashiwanoha, Kashiwa, Chiba, Japan*

[2]*Department Mechanical Engineering, Tokyo Institute of Technology, Ookayama, Tokyo, Japan*

[3]*Department of Information & Management Systems Engineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan*

Keywords: Memory, Audio Guide, Visual Behavior, Information Foraging Theory, Cognitive Model.

Abstract: Knowledge acquisition through appreciation behavior is a commonly experienced phenomenon. Appreciation behavior is characterized by real-time processing of information input through the five senses. This study focuses on multimodal information processing triggered during appreciation behavior, aiming to enhance knowledge acquisition, i.e., learning, by appropriately designing the provided information. While information during appreciation is presented as visual and auditory stimuli, learning is assumed to occur through the memorization of the content of the auditory information provided. Appreciation behavior is measured as visual behavior and modeled based on the well-established theory of information foraging. According to the information foraging theory, the process leading to information acquisition involves two states: the foraging state, where individuals actively seek information sources, and the acquisition transition state, where attention is directed towards information sources for acquiring information. Based on this theory, the characteristics of visual behavior are extracted for foraging and acquisition transition behaviors. This paper suggested that foraging state can be discerned by setting a threshold for gaze point movement frequency, while the acquisition transition state can be clearly delineated by examining the movement patterns of central and peripheral vision until reaching acquisition.

1 INTRODUCTION

Humans acquire and learn various information in the course of their lives. In the ambiguous and vast information society surrounding us, it is necessary to have our attention direct to the appropriate information to carry out the perceptual and cognitive processes for extracting useful information for learning. A method to support the smooth processing of this information is "multimodal information processing." The provision of information, which is designed by considering the characteristics of multimodal information processing, facilitates learning (Moreno and Mayer, 2007). In particular, the information tailored to the visual and auditory multimodal information processing facilitates smooth information processing and makes it easier to memorize information (Kitajima et al., 2019); this should promote learning effectively.

[a] https://orcid.org/0000-0001-9370-8443
[b] https://orcid.org/0000-0002-0310-2796

This study focuses on visual and auditory multimodal information processing in situations where an audio guide is provided alongside a video content. In a similar context, Hirabayashi et al. (2020) suggested an effective structure for memorization of video content, which consists of two parts; *visual guidance part (VG-part)* and *information addition part (IA-part)*. The VG-part is for directing the user's attention to a particular object in the scene by using superficial information such as its position, shape, and color. The IA-part is for providing supplementary information concerning the object indicated by the preceding VG-part, including its name and historical background. Hirabayashi et al. (2020) speculated on the reason why the presentation timing of VG-part and IA-part affects memorization of the contents of the movie. The specific presentation timing of these parts caused changes in visual behavior, which in turn caused changes in visual information processing to create a memory trace of the visual-audio experience.

Kurihara, Y., Shino, M., Nakahira, K. and Kitajima, M.
Visual Behavior Based on Information Foraging Theory Toward Designing of Auditory Information.
DOI: 10.5220/0012474600003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 530-537
ISBN: 978-989-758-679-8; ISSN: 2184-4321

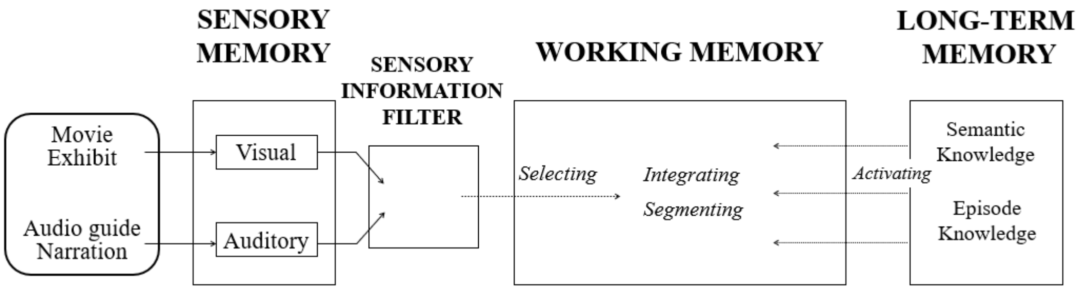


Figure 1: A cognitive model on memory formation based on Moreno and Mayer (2007).

However, the provided auditory information was pre-recorded, and real-time presentation timing tailored to individual is still unclear, leaving room for memory improvement. The purpose of this paper is to elucidate the effect of characteristics of visual behavior, which is real-time measurable, on memory by focusing on the information processing process. The purpose of this paper is to clarify the effect of the characteristics of visual behavior on memory by focusing on the information processing process.

This paper begins by organizing the memorization process for the input of multimodal information. Subsequently, the states of information processing leading to memorization of auditory information are considered based on information foraging theory (Pirolli and Card, 1999). By correlating the states with visual behavior during the provision of auditory information, the characteristics of visual behavior are highlighted. Following this, the characteristics of visual behavior during the provision of auditory information are elucidated, which indicate these states. Finally, the relationship between the characteristics of visual behavior in each state and the memorization of the provided auditory information is discussed.

2 EXTRACTION OF VISUAL BEHAVIOR FEATURES

For information to be memorized, it is necessary to consider how to form memories by relating representations obtained by perceiving stimulus input via sensory organs from the external environment to long-term memory. In the following subsections, a cognitive model is organized that shows memory formation through multimodal information processing targeted in this paper. Then, based on the information foraging theory, two states are assumed for the information process. Finally, features of visual behavior that discriminate between these states are extracted.

2.1 Memory Formation Through Visual Information Processing

Based on the cognitive-affective model proposed by Moreno and Mayer (2007), the process to obtain and memory information from the external environment through visual information is illustrated in Figure 1. The process unfolds as follows:

- **Perceptual Process:** Visual and auditory information is sent to the sensory memory via sensory organs and collected temporarily in each sensory memory. Only the information selected with the sensory information filter is further processed in working memory.
- **Cognitive Process:** Input information in working memory activates the knowledge associated with that information in long-term memory, and is processed together with the activated knowledge to be understood. Finally, the input information is integrated with existing knowledge and updates long-term memory.

In the case of learning with audio guide, auditory information is added to visual information processing, as illustrated in Figure 1. During this process, there might be cases where viewers do not pay attention to the auditory information in perceptual process. Similarly, if provided auditory information is not related to visual information, both information may not be integrated and may not be memorized in cognitive process. Therefore, it is essential to focus on visual behavior that reflects visual information processing and understand the effects of visual behavior on memory in response to the addition of auditory information.

2.2 Assumptions of the Information Processing Process Based on Information Foraging Theory

This study focuses on information foraging theory, which is a theory of information acquisition, to understand the effects of visual behavior on memory when auditory information is provided.

Information Foraging Theory (Pirolli and Card, 1999) involves that when exposed to an information source, the process is repeated: foraging behavior of information, then trying behavior to acquire information, and finally acquiring it. Foraging behavior, in this context, refers to looking around various information sources to evaluate their respective values. Trying behavior to acquire the information involves directing attention to information sources deemed valuable through foraging behavior and processing information related to those sources. Finally, acquisition signifies the successful incorporation of information into cognitive processes, leading to memory formation and comprehension.

How to evaluate information values in a foraging state is explained by the information scent model (Pirolli, 1997): Based on only superficial information, such as labels or icons of information sources, individuals anticipate the amount and content of information contained in the sources and assess how much desired information can be obtained.

Based on both theories, the process of information acquisition can be considered as a repeated sequence of two states and the acquisition of information.

- **Foraging State.** Foraging information sources that are likely to acquire the most effective acquisition of information based on information scent.
- **Acquisition Transition State.** Focusing on the chosen information source, trying to acquire the desired information.
- **Acquisition.** Acquiring desired information with high value.

Among the viewing conditions targeted in this study, it is considered that there is a foraging state where the participant searches for an appropriate information source among many objects and an acquisition transition state where the participant start trying to acquire that information source. Consequently, when auditory information is provided during the forage state, the participant is likely not to memory it because the participant is not yet in the acquisition state. Conversely, if the participant is in the acquisition transition state, the participants may or may not acquire it.

From the above, it is necessary to acquire information in the acquisition transition and not in the foraging state. Therefore, in the next chapter, visual characteristics that distinguish between the foraging and acquisition transition states concerning the memory of auditory information are extracted. In addition, the effects on memory will be discussed by clarifying the

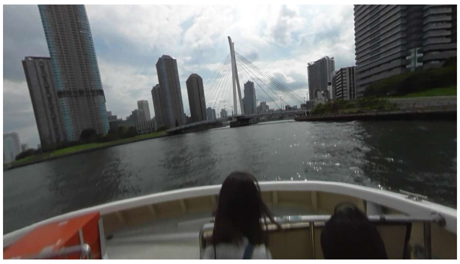


Figure 2: The movie used by Hirabayashi et al. (2020).

the characteristics of visual behavior leading up to acquisition in the acquisition transition state.

If the characteristics of visual behavior during the foraging and acquisition transition states can be clearly delineated, by avoiding the foraging state and aligning auditory information with the visual behavior characteristics leading to the acquisition during acquisition transition state, the memorization of auditory information can be enhanced.

2.3 Extraction of Visual Behavior Features Representing Two States

In this section, we extract the characteristics of visual behavior, which represent the foraging and acquisition transition states. To achieve this, visual behavior and memory are focused on, as measured in the prior research conducted by Hirabayashi et al. (2020), while providing audio guides during video appreciation. The video used in this research depicts the scenery visible from the front of cruise ship as if travels down the Sumida River (Figure 2). It features a slow flow in one direction, showcasing more than 10 objects. The audio guide consists of VG-part and IA-part, and provides commentary on *Paris square* appearing in the early part of the video and the the statue named *"Le Message"* (hereafter S_1) appearing towards the end, within the context of the friendly relationship between the Sumida River in Tokyo and the Seine River in Paris. The number of participants is 12(11 males and 1 female, average age = 21.67, SD = 0.62). Under these viewing conditions, visual behavior was measured using the Tobii Pro Glass 2. Memory evaluation was conducted through recall test.

Under these conditions, features of visual behavior that distinguish between a foraging state and an acquisition transition state are extracted by comparing visual behavior in the cases where content was memorized and those content is not memorized, based on the notion that the auditory information provided in a foraging state is not memorized. In cases where content was memorized, the visual behavior directed toward S_1 in the video is focused. In this scenario, a considerable number of participants shifted their gaze towards the target and its surroundings prompted by the VG-part, and there was a notable characteristic of the low frequency of gaze shifts. In cases where content was not memorized, the visual behavior directed toward "Paris Square" in the video is focused. In this scenario, the majority of participants didn't shift their gaze towards the target even after the presentation of the VG-part, and viewers directed their gaze towards various targets during the interval of IA-part. There was a notable characteristic of the high frequency of gaze shifts.

Based on the above, it can be inferred that, if participants successfully shifted their gaze to the explained target after the presentation of gaze-inducing segments, auditory information could be incorporated into information processing, depicted in Figure 1. Thus, when the "frequency of gaze shifts" is small, it can be inferred that auditory information is easily integrated into information processing. Conversely, when it is large, integration into information processing is considered to be challenging. In addition, another experiment with different participants under similar conditions showed that there was a negative correlation between the frequency of gaze shifts and memory (Kurihara et al., 2023). From the above, as distinguishing features of visual behavior for discerning between a foraging and an acquisition transition state, the "frequency of gaze shifts" is set. In the subsequent chapters, after distinguishing between a foraging state and an acquisition transition state based on the "frequency of gaze shifts", the study aims to elucidate the the characteristics of visual behavior leading to acquisition during acquisition transition states.

3 THE CHARACTERISTICS OF VISUAL BEHAVIOR IN FORAGING AND ACQUISITION TRANSITION STATES

In this section, we elucidate the characteristics of visual behavior based on the information foraging theory discussed in the preceding section. To achieve this, Section 3.1 provides the experiment to determine the threshold value for the frequency of gaze shifts that discriminates between the foraging and acquisition transition states, based on insights from the research conducted by Kurihara et al. (2023). Following that, Section 3.2 delves into clarifying the characteristics of visual behavior leading to acquisition during the acquisition transition state determined using that threshold value.

3.1 Discriminating Between the States

3.1.1 Experiment Overview

To discriminate between the foraging and acquisition transition states by the frequency of gaze shifts, it is necessary to calculate the frequency of gaze shifts in each state and set a threshold between them. To achieve this, participants were instructed with the three following task during video viewing. The first task is to forage for a target object presented before commencement, the second task is to press a button upon discovering the target object, and third task is to Memorize objects adjacent to the target object.

The first task makes participants to engage in an foraging activity, and it is presumed that it is a foraging state during this task. To facilitate this, we designed specific video and target selection conditions to ensure that foraging behavior towards the target is based on information scent. Under the video condition, the requirement was set for the number of objects in the video to exceed 10. Additionally, the target selection conditions dictated that the target should either occupy small area in the video screen or in the middle of the video presentation, which is about ten seconds after the start of the video. These conditions were considered to allow the participants to forage for the probable location or appearance of the target among numerous alternatives presented in the video, that is, the foraging state based on information scent.

The second task is conducted to capture the moment of transition between the first and third tasks. The third task instructs participants to memorize the surroundings of the target, and it is hypothesized that it is an acquisition transition state during this task. By calculating the frequency of gaze shifts during the first and third tasks separately, it is possible to establish appropriate thresholds for further analysis. The number of participants is fifteen (thirteen males and two females, average age = 22.60, SD = 1.20).

The procedure unfolded as follows: To have participants to practice the tasks, they watched some videos. Subsequently, participants engaged in the actual task of viewing eight videos as instructed. Finally, a questionnaire was administered to the participants regarding their prior knowledge of the images and the target objects.

3.1.2 Analysis Policy Based on Frequency of Gaze Shifts

To elucidate the characteristics of the frequency of gaze shifts during the foraging state, the number of gaze shifts from the start of the video was measured. The resulting graph, depicted in Figure 3, represents

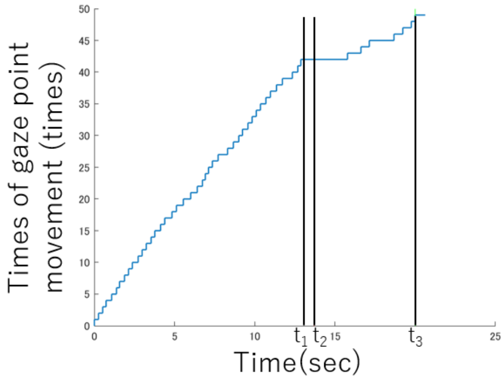


Figure 3: The example of the frequency of gaze shifts during appreciation (Kurihara et al., 2023).

the number of gaze movements on the vertical axis and time on the horizontal axis. In this graph, the frequency of gaze shifts is reflected as the slope, given its representation as the ratio of movement occurrences to time. t_1 denotes the time when the gaze first observed the target, t_2 denotes the time when the participant pressed the button, and t_3 denotes the time when objects other than those adjacent to the discovered target were first observed. Notably, if the condition for t_3 does not occur before the video ends, t_3 is set to 20 seconds, the end of the video. Following the methodology proposed by Kurihara et al. (2023), the foraging state encompasses the time from the beginning of the video to t_1. The frequency of gaze shifts during the foraging state is calculated by dividing the number of gaze movements that occurred during this time by the elapsed time. For the acquisition transition state, the time interval considered is from t_2 to t_3. The frequency of gaze shifts during the acquisition transition state is computed in a manner similar to that of the foraging state.

3.1.3 Result

The analysis results can be summarized as follows: As illustrated in Figure 3, the frequency of gaze shifts showed a higher tendency during tasks inducing the foraging state and a lower tendency during tasks inducing the acquisition transition state. The frequency of gaze shifts for each state was calculated for 15 participants, and 95% confidence intervals for the population mean were determined for each video. Consequently, thresholds were successfully set for four videos, as shown in Figure 4, created on the basis of Kurihara et al. (2023). A common threshold of 2.7 movements per second was established across the all videos. Therefore, foraging and acquisition transition states can be discriminated by focusing on the frequency of gaze shifts.

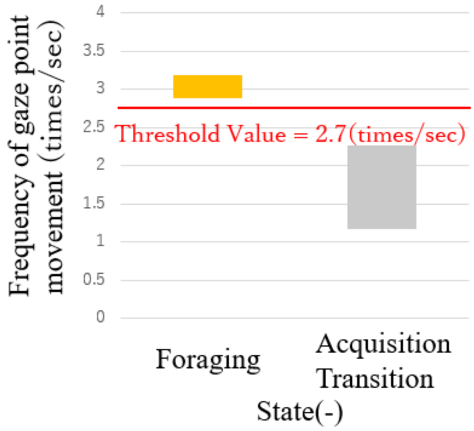


Figure 4: The establishment of threshold values.

3.2 The Acquisition Transition State

The threshold value obtained in the previous section allows the identification of the foraging state, where auditory information is not memorized. However, it is not clear whether the auditory information given in the acquisition transition state, where the frequency of gaze shifts is less than 2.7 times per second, contributes to memory. To clarify this, this section focuses on the characteristics of visual behavior leading to acquisition during the acquisition transition state. In particular, both central and peripheral vision are considered because as discussed in Section 2.3, participants were observed to gaze at the explanatory target and its surroundings during the provision of auditory information. In summary, this section aims to explore the characteristics of visual behavior leading to acquisition by focusing on how participants perceive the explanatory target through both central and peripheral vision during the provision of auditory information.

3.2.1 Subject of Analysis

The auditory information for the evaluation is the name of S_1, which is provided to all participants during periods when the frequency of eye movement is less than 2.7 times per second. The number of participants is twelve from the initial experiment as mentioned in Section 2.3, and an additional 8 participants from the supplementary experiment, making a total of 20 participants(seventeen males and three females, average age = 21.85, SD = 1.06).

3.2.2 Analysis Policy Focusing on Central and Peripheral Vision

This study conducts a temporal analysis of visual behavior along with auditory information for both cen-

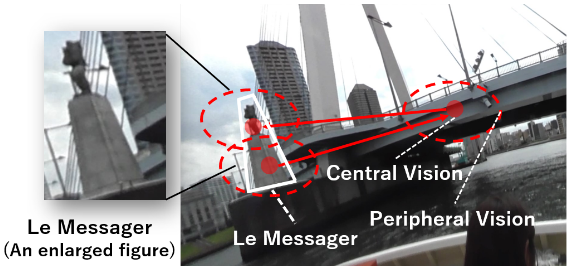


Figure 5: Visual behavior that removes the gaze from the explanatory object.

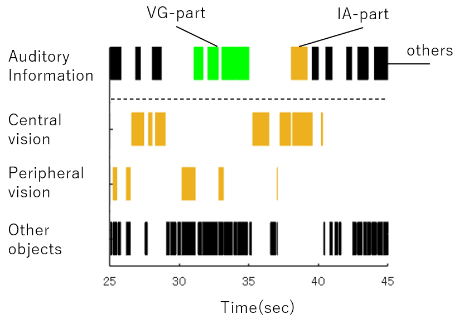
tral and peripheral vision. The gaze points were measured using the Tobii Pro Glasses 2. The recognition limit for symbols in peripheral vision is in the range of 5° to 30° from central vision (Yokomizo and Komatsubara, 2015). Moreover, it is understood that in areas with a wide field of view from central vision, detailed information processing becomes challenging, and processing by central vision is more sophisticated. Therefore, in this study, a field of view of 5° is set for peripheral vision.

The method of acquiring data for central and peripheral vision is shown based on Figure 5. The circle filled with red represents the central vision, while the oval with red frame represents the area with peripheral vision. First, regions are defined for each target along the boundaries of objects appearing in the video, as shown in the example of the S_1. For these regions, if central vision is within the designated area for a specific target X, it is considered that X has been captured through central vision. Furthermore, for peripheral vision, it is determined whether X was captured within a 5° field of view. Following these criteria, the duration is focused on which central and peripheral vision captured each target. Subsequently, an analysis is conducted to identify the characteristics of central and peripheral vision as part of visual behavior during the time when auditory information was provided.

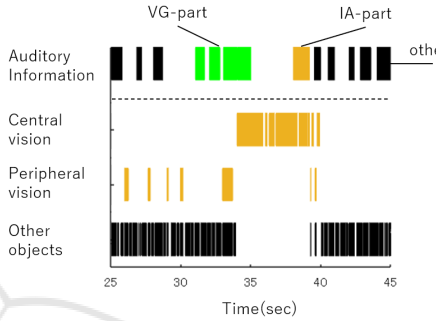
3.2.3 Results

Based on the analysis, the characteristics of visual behavior are aimed at clarifying when participants were able to retain auditory information during video viewing. Among the participants, three were able to remember the word "Le Message", which is the IA-part. These three participants are considered in the acquisition transition state.

Figures 6 depicts figures where the horizontal axis represents the elapsed time from the start of the video. The time range spans approximately 5 seconds before the explanation of the S_1 begins to about 5 seconds after the disappearance of the S_1 in the video. The



(a) Visual behavior of the participant who memorized the name of S_1 when auditory information was provided



(b) Visual behavior of the participant who did not memorize the name of S_1 when auditory information was provided

Figure 6: Visual behavior when auditory information was provided. (VG-part: Visual Guidance part, IA-part: Information Addition part).

uppermost part in each figure represents the auditory information. The green portion represents the VG-part to the S_1, the orange portion represents the IA-part where the name of S_1 is presented, and the black portion represents others. Figure 6(a) illustrates the chronological sequence of gaze behavior for participants who successfully memorized the S_1. The top section represents the time spent looking at the S_1 in central vision, the middle section represents the time spent looking at the S_1 in peripheral vision, and the bottom section represents the time spent looking at other objects in central vision besides the S_1.

As illustrated in Figure 6(a), it is evident from the participants' visual behavior that they redirected their gaze to a location where the S_1 could not be captured in peripheral vision. An example of this visual behavior is presented in Figure 5. First, the central vision captures the S_1, then the central vision moves to a place that could not be captured in the peripheral vision, and finally returned to the S_1 again. Moreover, by comparing the time when the central vision returned to S_1 and the time when IA-part was provided, as shown in Figure 6(a), it is evident that this occurred approximately 1 second before the onset of IA-part. However, as depicted in Figure 6(b), participants who consistently maintained central vision from the presentation of the VG-part until the presentation of the IA-part were not able to acquire the auditory information.

In summary, the following characteristics were identified for the visual behavior leading to acquisition during the acquisition transition state:

- Between the presentation of the VG-part and the IA-part, the explanatory target is captured in central or peripheral vision. Afterwards, the central vision is shifted to a place where the target cannot be captured in the peripheral vision, and then capture it again in the central or peripheral vision.
- Approximately 1 to 2 seconds before the presentation of IA-part, The explanatory target is captured in central vision.

Furthermore, for the participants in the acquisition transition state who were unable to acquire the information, they did not fulfill at least one of the characteristics.

3.3 Discussion

In this section, the visual behavior characteristics are discussed in the foraging state and acquisition transition state, as presented in Sections 3.1 and 3.2.

The foraging state was demonstrated to be identifiable by setting a threshold of 2.7 fixations per second for the frequency of gaze shifts. This 2.7 fixations per second corresponds to one fixation movement occurring approximately every 370 milliseconds. In the task instruction experiment conducted, participants were instructed to search for a pre-presented target, and it is reasonable to assume that they were engaged in shape matching, involving shifting their gaze. According to the Model Human Processor model (Card et al., 1986), a cognitive processing model, the sum of the time required for gaze movement (160 milliseconds) and shape matching (210 milliseconds) is 370 milliseconds, aligning closely with the threshold obtained in this paper. Thus, the threshold for the frequency of gaze shifts in the foraging state holds a similar significance from the perspective of the information processing process.

Subsequently, the two visual behavior characteristics, which are described in Section 3.2.3 regarding the acquisition transition state, are discussed. First, the characteristic of "looking at an object outside the peripheral vision area once" is considered based on the Optimal Foraging Theory. The Optimal Foraging Theory, a part of the information foraging theory (Pirolli and Card, 1999), is a theory related to the movement between multiple sources of information.

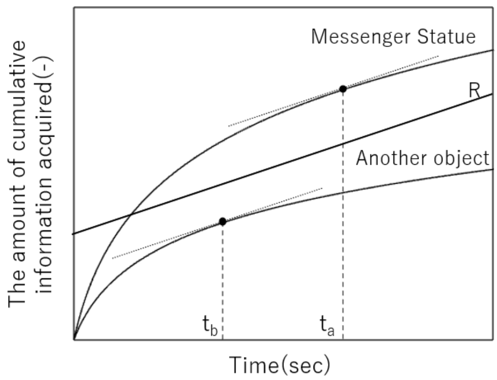


Figure 7: The relationship between information acquisition efficiency and time based on the optimal patch theory.

Specifically, the theory posits that the efficiency of information acquisition in a particular source decreases as one stays longer in that source, and movement between sources occurs when the efficiency falls below the average information acquisition efficiency.

Applying the Optimal Foraging Theory to the visual behavior during the appreciation of the target in this study, it can be visualized as shown in Figure 7. The horizontal axis represents the time spent attending to each target, and the vertical axis represents the cumulative information processed based on the duration spent on each target. R denotes the average information acquisition efficiency for all targets, while t_a and t_b represent the times at which the information acquisition efficiency for each target becomes equal to R. Based on this figure, the visual behavior obtained in Section 3.2 is correlated. Initially, attention is directed to the S_1 and information processing takes place. Subsequently, at the time t_a when the information acquisition efficiency within that source decreases below the average information acquisition efficiency R, gaze is shifted to another target to enhance the efficiency of information acquisition. Then, by redirecting gaze back to the S_1, the information acquisition efficiency for the S_1 increases again, and in this state, the word *"Le Message"*, which is the IA-part, is assigned.

Therefore, the visual behavior that diverts gaze away and then returns it to the explanatory target is considered a behavior aimed at increasing the efficiency of information acquisition. Thus, it reflects an attempt to enhance the activity related to the explanatory target through auditory information, and it implies that information is acquired in a state of heightened activity related to the explanatory target when auditory information was provided. The second characteristic of visual behavior is considered. It was observed that when auditory information was provided immediately after viewing the S_1, it was less likely to be remembered. This result aligns with the insights of Hirabayashi et al. (2020), suggesting that when auditory information is provided shortly after observing an explanatory target, the increased information processing load makes it less memorable. Therefore, to enhance the memorization of IA-part, it is necessary to have the explanatory target in view several seconds before the timing of information addition.

In summary, the obtained characteristics of visual behavior are considered to reflect the states during information processing in the foraging and acquisition transition phases.

4 CONCLUSIONS

In the information processing with auditory information presentation during appreciation, applying the information foraging theory and establishing foraging and acquisition transition phases allowed us to grasp the characteristics of visual behavior. Consequently, the following insights were obtained:

- By considering the frequency of gaze shifts, it is possible to identify a foraging state in which auditory information cannot be processed, specifically when the frequency of gaze shifts is above 2.7 times per second.
- The visual behavior characteristic of "diverting gaze from the explanatory target after the presentation of gaze-inducing cues, then returning to the explanatory target with central vision before the presentation of additional information" allows the acquisition of auditory information in the acquisition transition state.

Based on the identified visual behavior characteristics, designing auditory information for enhanced memory retention can be achieved through the following steps: Initially, provide viewers with auditory information during the acquisition transition state, determined based on frequency of gaze shifts. Subsequently, the visual behavior is focused on, which is related to the explanatory target, to determine whether auditory information has been acquired. In cases where it is determined through these visual behaviors that the information has not been acquired, appropriate measures, such as reapplying auditory information based on gaze behavior characteristics, can be taken.

In the future, we will investigate whether auditory information during the viewing designed as described above improves memorization.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334, and National University Management Reform Promotion Project.

REFERENCES

Card, S. K., Moran, T. P., and Newell, A. (1986). The model human processor: An engineering model of human performance. *Handbook of perception and human performance*, (45):1–35.

Hirabayashi, R., Shino, M., Nakahira, K. T., and Kitajima, M. (2020). How auditory information presentation timings affect memory when watching omnidirectional movie with audio guide. In *Proceedings of VISIGRAPP 2020, Vol. 2: HUCAPP*, pages 162–169. INSTICC, SciTePress.

Kitajima, M., Dinet, J., and Toyota, M. (2019). Multimodal interactions viewed as dual process on multidimensional memory frames under weak synchronization. In *COGNITIVE 2019 : The Eleventh International Conference on Advanced Cognitive Technologies and Applications*, pages 44–51.

Kurihara, Y., Motoki, S., Nakahira, K, T., and Kitajima, M. (2023). An Analysis of Eye Movements during a Visual Task (in Japanese). In *Proceedings of the 22nd Forum on Information Technology*, volume 2023, pages 95–98.

Moreno, R. and Mayer, R. (2007). Interactive multimodal learning environments. *Educ Psychol Rev*, 19:309–326.

Pirolli, P. (1997). Computational models of information scent-following in a very large browsable text collection. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 3–10, New York, NY, USA. Association for Computing Machinery.

Pirolli, P. and Card, S. (1999). Information Foraging. *Psychological Review*, 106:643–675.

Yokomizo, K. and Komatsubara, M. (2015). *Ergonomics for Engineers, Revised 5th edition (in Japanese)*. Japan Publication Service.