How Auditory Information Presentation Timings Affect Memory When Watching Omnidirectional Movie with Audio Guide

Rinki Hirabayashi[1], Motoki Shino[1], Katsuko Nakahira T.[2] and Muneo Kitajima[2]
[1]*Department of Human & Engineered Environmental Studies, The University of Tokyo, Kashiwanoha, Kashiwa, Chiba, Japan*
[2]*Department of Information & Management Systems Engineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan*

Keywords: Memory, Audio Guide, Omnidirectional Watching, Information Acquisition, Cognitive Model.

Abstract: This study focuses on audio guide as a support for smooth information acquisition for visual stimuli. The interval between provision timing of visual guidance part, which explains explicit features of the object, and information addition part, which explains implicit features of the object, is set as a parameter and its effect on memory is measured as an indicator for estimating the degree of smoothness in information acquisition. Eye tracking experiments were conducted in a dome theater with the omnidirectional movie using three timing interval conditions: shorter than two seconds (Short Interval), longer than three seconds and shorter than five seconds (Medium Interval), and longer than six seconds (Long Interval). The results showed that the memory scores for the movie presented in the Medium Interval condition was the largest. This paper discusses how the presentation in the Medium Interval condition allowed effective integration of visual information and the auditory information provided by audio guide: the visual guidance part of audio guide helped the viewer to find the objects at the best timing before the presentation of information addition part. This would have enabled the participants to elaborate the visual scene with the relevant long-term memory for integration with the auditory information.

1 INTRODUCTION

Information acquisition is one of the most difficult activities in life. This is because the information in the environment is ambiguous. In order to acquire necessary information, we need to pay attention to a right object in the environment and memorize it in the current context by performing a cyclic loop of perception and cognition process smoothly. A lot of supports are available to make this smooth information acquisition possible. One of them is the use of multimodal presentation of information. The effect of multimodal information on human behavior is discussed in such areas as multimodal learning (Moreno and Mayer, 2007) and cognitive science (Kitajima et al., 2019). To extend this line of research, it is effective to apply the method of support, "multimodal presentation of information", to a variety of environments, and identify defining characteristics that should result on effective support for information acquisition.

As the first step for it, this paper applied multimodal information presentation to the omnidirectional movie with audio guide. In the study that examined the characteristics of information acquisition while watching omnidirectional movies with audio guide (Kitajima et al., 2017), the presentation timing of audio guide, which was either *before* or *during* watching the movie, was identified as one of the important features that should affect memorization of the contents of the movie. This paper focuses on *contents of the audio guide* and *its presentation timing* as well, which are pointed out as features of the audio part of movie (Kanai, 2000), including situation, timing, and contents.

The purpose of this paper is to clarify the effect of audio guide presentation timing on memory when watching the omnidirectional movie. This paper begins with the section to discuss cognitive framework that shows the effect of audio guide timing on memory and related work on making experience memorable. The following section describes an eye tracking experiment using a dome theater environment. Eye movements were recorded to know the acquisition timing of visual information and this was used to discuss the multimodal interaction of visual and auditory acquisition timing. Finally, the last sections present

Hirabayashi, R., Shino, M., T., K. and Kitajima, M.
How Auditory Information Presentation Timings Affect Memory When Watching Omnidirectional Movie with Audio Guide.
DOI: 10.5220/0008966201620169
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 2: HUCAPP*, pages 162-169
ISBN: 978-989-758-402-2; ISSN: 2184-4321

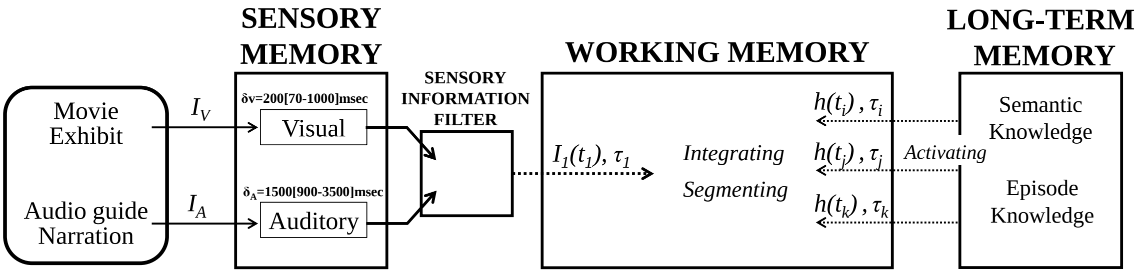


Figure 1: A cognitive model on memory formation.

the experimental results and discuss how timings of audio guides affect memory.

2 IMPORTANCE OF TIMING ON EFFECT OF AUDIO GUIDE

To make experience memorable, it is essential to consider how memory is formed based on the perceived stimulus that comes from external and internal environment. Starting from the introduction of cognitive model of memory formation, this section discuss why it is essential to take timings into account.

2.1 Memory Formation When using Audio Guide

Memory formation starts with acquisition of information from the internal and external environment. For acquiring visual information with the support of an audio guide, consideration of multimodal information acquisition is essential. Using Moreno's cognitive-affective model (Moreno and Mayer, 2007) as a base, a simplified cognitive model is derived as shown by Figure 1. I_v and I_A each represents the acquired visual and auditory information. δ_V and δ_A each represents the half-life period of visual and auditory information in sensory memory. $h(t_i)$, $h(t_j)$, and $h(t_k)$ represent the activated portions of knowledge initiated by I_1 at different timings. t_i, t_j, and t_k represent the timings when portions of long-term memory are retrieved for inclusion in working memory. τ_i, τ_j, and τ_k represent the durations that the respective information, $h(t_i), h(t_j)$, and $h(t_k)$, are present in working memory.

The process of acquiring knowledge from visual stimuli and audio guide goes as follows (Kitajima and Toyota, 2015):

1. **Perceptual Process**
Information, such as visual appearance and audio guide associated with it, is collected via sensory organs. Acquired information is temporarily stored in individual sensory memories for a short period of time. When the information passes through Sensory Information Filter, only a fragment of information is selected for further processes in working memory. Pieces of information originated from different modalities are processed independently and in parallel.

2. **Cognitive Process**
Inputted information in working memory works as a trigger for activating related knowledge from long-term memory. The inputted information is mentally organized with the activated knowledge and other related information. Through this process, our brain makes sense of the inputted information. Finally, inputted information is integrated with existing knowledge and form a new memory.

Through this process, the acquired information establishes links with the existing memory networks and as a result, it is memorized. Therefore, not only the acquired information itself but also the other pieces of information that is available at the same time plays an important role in how memorable the acquired information is. This is where timing comes in. Even if the same pieces of information are presented, if the pieces of information are perceived in different timings, it directly affects the quantity of information available for integration and organization of the acquired information. This timings' effect on the quantity of information is shown in Figure 2. Depending on the different perception timings, the quantities of information in working memory can be classified into three types. In this study, we call these types Mode 1, 2, and 3. For the sake of simplicity, let's consider a situation where information is perceived in two different timings, t_1 and t_2. h_i and h_j denote the activated portions of long-term memory triggered by the first information, and h_k and h_l denote the activated portions of long-term memory triggered by the sec-

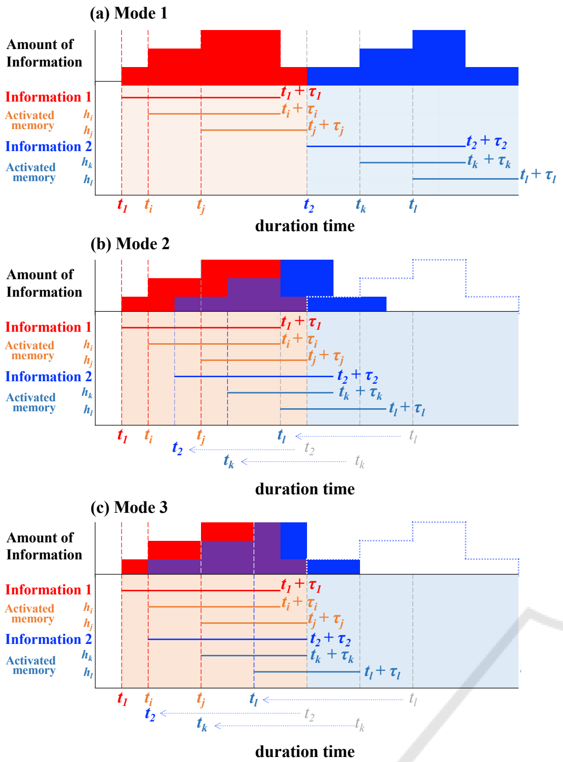


Figure 2: Timeline of three Modes.

ond information. t_i, t_j, t_k, and t_l represent the timings when portions of long-term memory, h_i, h_j, h_k, and h_l, are retrieved from long-term memory for inclusion in working memory.

The solid horizontal lines in Figure 2 represent the duration that the respective pieces of information, h_i, h_j, h_k, and h_l, are present in working memory. In addition, there are two solid horizontal lines that indicate the periods of the first and second information present in working memory. τ_i, τ_j, τ_k, τ_l, τ_1, and τ_2 are defined by the starting times and the ending times of the respective sources of information. For a specific time during the period of $[t_1, t_l + \tau_l]$, the amount of information available is expressed by the horizontal lines that exists. A histogram above the timeline shows the total amount of information present in working memory at a specific time. The histograms shaded in red represent the information triggered by the first information and those shaded in blue represent the information triggered by the second information. The overlapped areas of the red and the blue shaded histograms are shown in purple. All the three modes have the same timing for the first information, but each mode has different timings for the second information. Therefore, the red shaded histogram stays at the same place, while the blue shaded ones appear in different positions: the distance between the red and the blue shaded histograms are the farthest in Mode 1 and the closest in Mode 3.

In order to define the modes unambiguously, the following conditions are imposed for the starting times of information 1, 2 and the respective activated portions of long-term memory: $t_1 < t_2$, $t_1 < t_i < t_j$, and $t_2 < t_k < t_l$. Each mode is defined as follows:

- **Mode 1 (No Overlap):** This mode is shown in Figure 2 (a) and the condition for this mode is expressed as $t_j + \tau_j < t_2$. In this mode, the second information is perceived after the disappearance of the activated portions of long-term memory triggered by the first information. Therefore, the two pieces of information are integrated independently. In this mode, no overlap exists for the two sources of information.
- **Mode 2 (Moderate Overlap):** This mode is shown in Figure 2 (b) and the condition for this mode is expressed as $t_i < t_2 < t_j$. In this mode, the second information is perceived before the disappearance of the activated portions of long-term memory triggered by the first information. Therefore, the two pieces of information are present at the overlapped times and they are available for the cognitive processes that proceed. In this mode, the size of the area of the overlapping pieces of information, which is shown in the purple shaded histogram, is moderate.
- **Mode 3 (Large Overlap):** This mode is shown in Figure 2 (c) and the condition for this mode is expressed as $t_2 < t_i < t_j$. In this mode, the second information is perceived before the appearance of the activated portions of long-term memory triggered by the first information. Therefore, similar to Mode 2, the two pieces of information are present at the overlapped times and they are available for the cognitive processes that proceed. In this mode, the size of the area of the overlapping pieces of information, which is shown in the purple shaded histogram, is large.

In order to make the full use of the overlapped area of the histograms, it is essential to consider the relationship between cognitive ability and the amount of information available for integration. If the amount of information is smaller than the amount our cognitive ability can work out, it can establish more links with the existing memory network by additional information inputs. This can be seen in phenomenon like modality effect (Mayer, 2002), where more memorable education was done using more than one modality to input more pieces of information simultaneously. On the other hand, if the amount of information is larger than the amount our cognitive ability can work out, inputting more information can cause cognitive overload and do harm, like reverse modality effect (Tabbers et al., 2004). In sum, what mode a viewer is in has effect on memory and the timing is the major factor for it.

2.2 Effect of Audio Guide Timing When Watching Movie

To make experience memorable with audio guide, it is essential to consider perception timings of two modalities, visual stimuli and audio guide. In our previous study (Hirabayashi et al., 2019), to examine an audio guide timing, an audio guide was decomposed into two parts based on its contents, 'visual guidance part' and 'information addition part'. *Information addition part* is the part of audio guide that provides explanation about the objects that are not visible, such as the names and impressions of the objects. Information presented in this part would work while integrating the gist of the explained subject with the visually acquired information. *Visual guidance part* is the part of audio guide that concerns visible and easy to recognize entities, such as places and color. They would not be a good subject for integration but explain what and where the objects explained in 'information addition part' are and can be used as a guidance to find them. As a standard construction of audio guide, 'visual guidance part' is presented first and 'information addition part' follows to explain the object that the viewer's attention should have already been guided to it.

Our previous study (Hirabayashi et al., 2019) focused on the relationships between the time when a specific information addition part is provided and the time when the participant finds the object that is explained by it.[1] In order to detect the time when the participant begins to pay attention to a specific object presented on a 2D display, eye movements were classified into two types: (1) voluntary eye movements to perceive visual information to be further processed by cognitive processes, and (2) involuntary eye movements, not for further cognitive processes. Voluntary and involuntary eye movements were identified according to the criteria reported by Ohtani (Ohtani, 1971). An eye movement whose fixation time is longer than 270 msec is judged as voluntary and the one shorter than 270 msec as involuntary.[2] The objects that are fixated with voluntary eye movements are considered as those to be processed by the cognitive processes that follow.

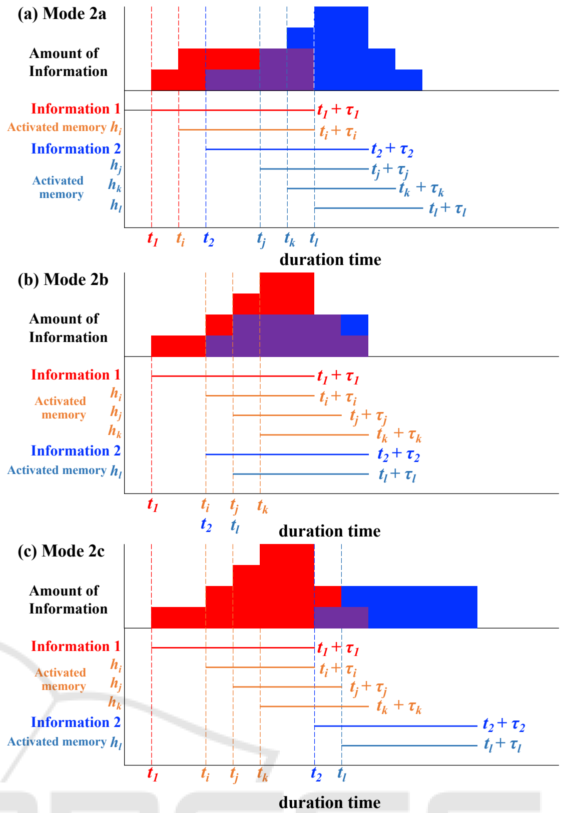


Figure 3: Timeline of three difference cases of Mode 2 in Figure 2.

The study (Hirabayashi et al., 2019) found that the participants who listened to the information addition part of the audio guide *after* visually finding the object memorized it better than those who listened it *before* visually finding it. The situations that were analyzed in the experiment are as follows: the participant fixates the object visually and activates its associated memory for inclusion in working memory, and he or she listens to the information addition part of audio guide and activates its associated memory for inclusion in working memory. The intervals between visual guidance part and information addition part were *not* manipulated in this experiment but it was possible to find situations where two pieces of information from each modalities are perceived at the timing that are characterizes as Mode 2 shown in Figure 2. The result gained in the study (Hirabayashi et al., 2019) showed that Mode 2 has to be treated at a finer grain

[1]Stimuli and evaluation used in the study (Hirabayashi et al., 2019) were identical in nature with those used in the experiment stated in Section 3. Major difference is that our previous study used 2D display and 2D movie instead of 3D dome theater and omnidirectional movie.

[2]Interval between saccadic movement was plotted to the Weibull chart and show that there is a turning point at 270 msec. Through the visual task analysis, a fixation with the fixation time less than 270 msec was classified as an involuntary movement and fixation with fixation time longer than 270 msec, was classified as a voluntary movement.

size considering the processes that should follow 'after finding the object that is to be explained by the audio guide' and 'after listening to the information addition part of audio guide'. Figure 3 shows three submodes in Mode 2. The differences in the amounts of knowledge activated by the inputted information are taken into account. The area of histogram is smaller for the case where input information is auditory, and the area of histogram is larger for the case where input information is visual. Mode 2a shows the case where the first information is auditory, and Mode 2b shows the case where the first information is visual. Mode 2c shows the case where the first information is visual, but the second auditory information is perceived in delay.

The participants who listened to the information addition part of the audio guide *after* visually finding the object was in Mode 2b and those listened to the information addition part of audio guide *before* visually finding it were in Mode 2a. As a result, the former formed a better memory than the latter.

Thus, to design the audio guide, it is essential to consider the interaction between visual finding timing and additional information provision timing. In this paper, the interval between *visual guidance part* and *information addition part* is manipulated and effectiveness of using this interval as the parameter for designing audio guide is discussed. Eye tracking experiment was conducted in a dome theater with the omnidirectional movie, which simulated the real environment in terms of a wide field of view that allows the viewer to appreciate a large amount of visual information.

3 METHOD

An experiment was conducted to understand the relationship between timing interval, which is the interval between visual guidance part and information addition part of the audio guide, and memory in the omnidirectional environment. Since the effect of audio guide works in relation with the viewing behavior, eye tracking was conducted during the experiment. Note that, in the description of the experiment, the expression "target object" or "target" refers to the object explained in the audio guide.

Eight participants (all males, average age = 23.25, SD = 2.44) took part in the eye tracking experiment and no one had visual or health problem on taking the experiment.

3.1 Conditions

Three conditions were considered for timing interval, denoted as $\hat{T}$ hereafter, as follows:

- Short interval condition (SI), $0 \leq \hat{T} \leq 2$ [sec]: In this condition, information addition part of the audio guide is likely to be presented before the participant found the target object. The results of our previous study predict that the information presented in this condition should be less memorable.
- Medium interval condition (MI), $3 \leq \hat{T} \leq 5$ [sec]: In this condition, information addition part of the audio guide is likely to be presented after the participant found the target object. The results of our previous study predict that the information presented in this condition should be more memorable.
- Long interval condition (LI), $\hat{T} \geq 6$ [sec]: Like the MI condition, information addition part of the audio guide is likely to be presented after the participant found the target object. The results of our previous study predict that the information presented in this condition should be more memorable for only a limited number of objects.

3.2 Stimuli

Three omnidirectional movies were prepared. Each movie had a respective audio guide with one of the three conditions. The movies showed the landscape taken from a slow-paced boat going down the Sumidagawa River in Tokyo. A movie taken from a slow-paced boat was chosen as a stimulus for this experiment because it is likely to contain scenes or targets that satisfy the conditions:

1. The target in the movie should move in a slow pace. This condition is needed to make the target appear and stay in the field of view long enough for a viewer to take needed visual information of the target object.
2. The target should not be easy to notice without a guidance. This condition is needed to refrain viewers from paying attention to the target beforehand and to see the effect of audio guide clearly.
3. Scenes should contain many objects to look at throughout the movie. This condition is needed to simulate the situations where audio guide is in need.

Since the position of a target in the omnidirectional environment was difficult to indicate, a visual marker was superimposed to the original movie to point the target. A visual marker appears after the

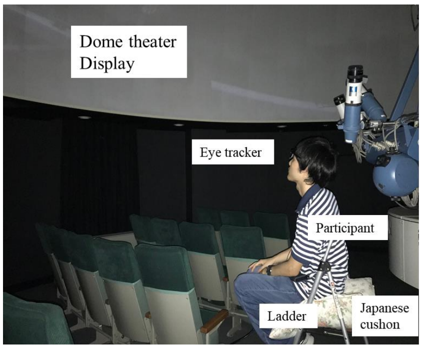


Figure 4: Experiment environment.

presentation of visual guidance part of audio guide and remains for two seconds.

3.3 Apparatus

The stimuli were controlled and projected to a dome theater by using StellaDome Professional (AstraArts, Shibuya), which is 8 meters in diameter and 2 meters off the ground, located on Teganuma Aquatic Park “Mizunoyakata”. Viewing behavior was recorded using a wearable eye tracker (Tobii Pro Glasses 2) at a sampling rate of 50Hz. Eye positions and gaze points were calculated with the 3D eye model and gaze mapping algorithms. These gaze points were recorded as a coordinate on the video taken from the scene camera. This scene camera has resolution of 1920×1080 and covers 82 degrees horizontally and 56 degrees vertically.

The experiment was carried out with one participant at a time. The participants were seated on a ladder and their head positions were located approximately 3.5 meters from the nearest point on the dome theater and approximately 1.6 meters from the ground. Figure 4 shows the arrangement of the experiment.

3.4 Evaluation

In this study, a questionnaire was conducted to investigate information that participants memorized when viewing movies. Since the questionnaire was conducted soon after viewing the movie, a recall test was chosen to see the differences explicitly. To make the quantitative evaluation of memory, the replies from the participants were broken down to meaningful units using a morphological analysis technique. They were scored from the quantitative and qualitative perspectives, by giving points to those units that are related to the targets and the narrative contents spoken in the audio guides (one point was given to a noun or a verb, two points to a pronoun). Those points were summed up to define the Memory Score. For example, if the participant answered ‘Statue of Messenger was given for its friendship’ in the questionnaire, it is broken down to ‘Statue of Messenger’, ‘given’, and ‘friendship’. Since the narration was ‘The wooden replica of Statue of Messenger was given to Tokyo as a proof of frienship...’, each broken down unit was given points, i.e. two points, one point, one point, and as a total, four points were given as the Memory Score.

3.5 Procedure

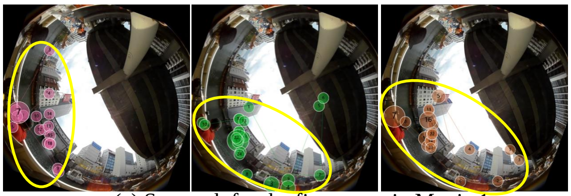
Before viewing the movie, participants were told to make themselves comfortable and to view the movie freely in order to simulate the actual viewing behavior.

During the experiment, six movies were presented to the participants. Three movies (Movie 1, 2, 3) were for analysis, two movies (Movie A, B) were for dummy, and one was for instruction. Movie 1, 2, and 3 are the stimuli explained in Section 3.2 and used to examine the effect of timing intervals on memory. Movie A and B were used as the first and the last movie. These two movies were used to eliminate the known effect that the first and last movies were memorable regardless of the effects of timing intervals. An instruction movie was used to explain how the directions of the targets are described using the clock positions in the visual guidance part of audio guide. It was presented after Movie A or B and before Movie 1, 2, and 3. Each movie was approximately 1 minute long, and 10 seconds intervals were inserted between the movies. Considering the order effect, each participant was presented with a randomized patterns. After the participants finished viewing the movies, they were asked to complete the questionnaire.

4 RESULT

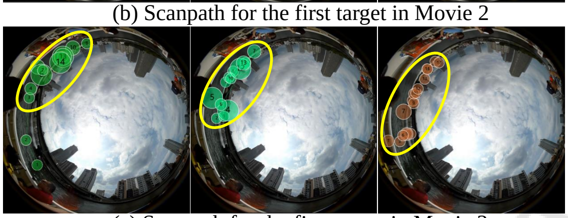
An analysis was conducted on the basis of the objects presented in the movies and six target objects, two from each movie, were chosen as the candidate target objects for further analysis. One of the candidates was found not to be suitable for the analysis because the lengths of the time it appeared in the field of view of the participants were not long enough for the LI condition being valid (4.9 seconds in average (SD = 1.3)). This target object was removed from the analysis.

To understand the effect of audio guide on memory, viewing behaviors were analyzed. In this section, scanpath and visual finding timings were examined.



(a) Scanpath for the first target in Movie 1

(b) Scanpath for the first target in Movie 2



(c) Scanpath for the first target in Movie 3

Figure 5: Scanpath for the first target in three movies.

The distinctive scanpath seen in omnidirectional movie environment was observed for the first target of the movie. Figure 5 (a), (b), and (c) show the scanpaths for the first targets of the respective movies from the start of the movie to the start of information addition part of audio guide of the first target. The eye movement data with the largest number of fixation points is chosen to represent the characteristic scanpath for each target.

The scanpath that was seen distinctively in the omnidirectional movie environment is shown in Figure 5(a) and Figure 5(c). As shown in the figures, the participants looked around the dome theater to grasp the surroundings when the movie started. This was a typical scanpath for the omnidirectional movie environment and rarely seen on 2D movie (Hirabayashi et al., 2019). This shows that looking around type of scanpath was needed for omnidirectional movie due to its broad display area.

On the other hand, there were a few instances without a looking around scanpath as shown by Figure 5(b). This kind of scanpath was especially observed in this target. This target had a distinctive characteristic that the other targets didn't have: its visual guidance marker was provided only 1.6 seconds after the movie started. This kind of viewing behavior was expected to occur by the very short interval between the start of the movie and the start of the visual guidance part of the audio guide. Since it is reasonable to consider that the reactions of the participants to this particular target should have been different from those for the other targets, this target was removed from the analysis.

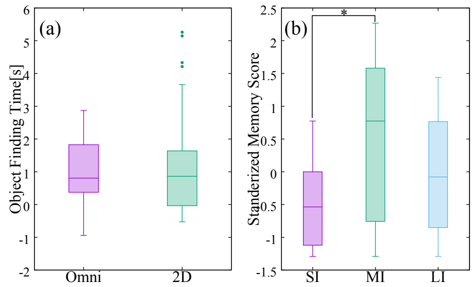


Figure 6: Results.

Object finding timing is shown in Figure 6(a). To evaluate the object finding timing, 2D movie data from our previous study is presented for comparison. Thirty-one samples were used for the omnidirectional movie group and fifty-three samples were used for the 2D movie group. Object finding timing is the time it took to find the target object after listening to the visual guidance part of the audio guide. The omnidirectional movie group took 0.94 seconds in average to find the target object (SD = 0.90). The 2D movie group took 1.20 seconds in average to find the target object (SD = 1.51).

Relationship between timing intervals of the audio guide and the Memory Score are shown in Figure 6(b). To exclude the effect of target differences, the Memory Scores are standardized for each target. The sample sizes for the respective conditions are ten for the SI condition, ten for the MI condition, and twelve for the LI condition. Using Mann-Whitney U test, a statistically significant difference was seen between the SI condition and the MI condition. Although there was no statistically significant difference between the MI condition and the LI condition, the tendency that the MI condition scored higher than the LI condition did was observed.

5 DISCUSSION

The MI condition, which has the timing interval longer than three seconds and shorter than five seconds, resulted to be the most memorable compared to the other timing conditions. As stated in Section 2.2, the MI condition resulted in the most memorable because this timing interval probably induced participants to acquire information in Mode 2b.

The effect of the viewing behavior on this result can be interpreted using the result of the object finding timings shown by Figure 6(a). Irrespective of the viewing conditions, most of the participants found the target object at approximately equal timings that favored the MI condition and there were only few who found the target object in shorter or longer timings. For the participants who took longer times to find the target object, the LI condition would be the timing interval that should have induced Mode 2b instead of Mode 2c. On the other hand, for the participants who took shorter times to find the target object, the SI condition would be the timing interval that should have induced Mode 2b instead of Mode 2a. Since the omnidirectional movie experiment resulted in smaller variance in object finding timing compared to the 2D movie experiment as shown by Figure 6 (a), it probably restrained such experimental trials that required longer or shorter times in object finding. Therefore, participants obtained the visual information at the similar timing and for those participants, the MI condition was ideal timing for presenting information addition part.

This result on viewing behavior can be further discussed by the presence of visual marker. This made it easy for the visual system to detect the targets and induce a quick response. It effectively decreased the number of participants who took long times to locate the target object. The necessity of visual guidance part of audio guide and visual marker is to help the participants quickly find the target object and works for aligning the time to start the timer to begin the information addition part of audio guide. In this timer aligned situation, the timing of the visual marker to appear should have worked for the restraint on the end of the timer when the information addition part to start. This restraint should have increased the number of participants who could synchronize audio with vision smoothly in the MI condition. Smooth synchronization and integration of multi-modal information work positively on memorization, as it is stated on theorical simulation conducted in the previous study (Kitajima et al., 2019). As a result, the MI condition marked remarkable effect on memory.

6 CONCLUSION AND FUTURE WORK

This paper investigated the effect of the timing interval of audio guide on memory. A model of modes based on timing differences was introduced to make rational explanations for relationship between timing intervals and memory. The omnidirectional movie experiment showed that timing intervals can be used to make experience memorable in omnidirectional environment.

For now, this study only focsed on the omnidirectional movie viewing behavior to simulate the real environment. But, for applying it to our everyday life, such as museum, gallery, guided tours, etc., it is important to apply and examine what this paper found in the outside experiment situations. Also, since the viewing behavior plays a significant role, there is need to examine the effect in movies with different characteristics like the fast moving targets or immovable targets.

ACKNOWLEDGEMENTS

We would like to express our appreciation to Abiko City Office for letting us use the dome theater for this experiment. This work was partly supported by JSPS KAKENHI Grant Number 19K12246.

REFERENCES

Hirabayashi, R., Motoki, S., Nakahira, K, T., and Kitajima, M. (2019). How auditory timing affect memory when watching movie with audio guide. *FIT2019*, 3:225–228.

Kanai, A. (2000). Rhetoric and viewpoint in film cognition. *Cognitive Studies*, 7(2):172–180.

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

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

Kitajima, M. and Toyota, M. (2015). Multi-dimensional memory frames and action generation in the mhp/rt cognitive architecture. *Procedia Computer Science*, 71:202–207.

Mayer, R. E. (2002). Multimedia learning. *The Annual Report of Educational Psychology in Japan*, 41:85 – 139.

Moreno, R. and Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19:309–326.

Ohtani, A. (1971). An analysis of eye movements during a visual task. *Ergonomics*, 14(1):167–174.

Tabbers, H. K., Martens, R. L., and Van Merrienboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, 74(1):71–81.