Can Pupillary Responses while Listening to Short Sentences Containing Emotion Induction Words Explain the Effects on Sentence Memory?

Shunsuke Moriya[1], Katsuko T. Nakahira[1][a], Munenori Harada[1], Motoki Shino[2] and Muneo Kitajima[1][b]

[1]*Nagaoka University of Technology, Nagaoka, Niigata, Japan*

[2]*The University of Tokyo, Kashiwa, Chiba, Japan*

Keywords: Emotion Induction Word, Pupil Dilation Response, Memory, Contents Design.

Abstract: In content viewing activities, such as movies and paintings, it is important to retain and utilize the viewing experience in memory. We have been studying the effect of the content of visual and auditory information provided during viewing activities and presentation timing on content memory. We have clarified the appropriate timing of presenting visual information that should be supplemented by auditory information. We have also found that the inclusion of emotion induction words in the auditory information is effective in forming content memory. In this study, we present a framework for examining the effects of emotion-evoking characteristics of short sentences while taking into account individual differences in memory. Subjects were presented with a short sentence with an emotion-inducing word at the beginning of the sentence, in which the impression of the entire short sentence would appear at the end of the sentence. We designed an experimental system to clarify the relationship between subject-specific pupillary responses to the emotion induction words and memory for short sentences. Our findings indicate a scheme that relates the pupillary response to short sentence memory.

1 INTRODUCTION

The development of digital technology has enabled the dispensing of knowledge by combining various types of digital content. There are two types of knowledge: explicit knowledge, which can be explicitly expressed by symbols, and latent knowledge, which is understood by reading between the lines. Digital technology is suitable for expressing explicit knowledge because of its high affinity to symbolic representation. On the other hand, to express latent knowledge through digital content, it is necessary to understand how humans, the recipients of the information, process digital information.

We have focused on viewing behavior as an example of latent knowledge transfer, including experience, and have studied the effects of the content and timing of presenting visual and auditory information during viewing behavior on the memory of contents in terms of both quantity and quality of information provided. With regard to the quantity of information, according to Hirabayashi et al., when there are two-tracks of information sources (e.g., visual and auditory), I_1 and I_2, to be related, the information can be retained without information overload by presenting I_1 and I_2 with a certain time interval between them (Hirabayashi et al., 2020).

Regarding the quality of information, Murakami et al. conducted an experiment on memory with a particular focus on auditory information. They created explanatory text, including emotion induction words, played auditory stimuli with the explanatory text read aloud along with the video, and had the participants listen to the video. The results of the impression evaluation and replay test of the video content showed that the participants' memory of the video was not affected by the stimuli (Murakami et al., 2021).

The results of these studies suggest that the depth of the learner's memory, i.e., the acquisition of lateral knowledge, is strongly connected with the learner's emotions and information about the surrounding environment. Therefore, it is important to design learning contents for the transfer of lateral knowledge by generating emotion and including the surrounding environment.

[a] https://orcid.org/0000-0001-9370-8443

[b] https://orcid.org/0000-0002-0310-2796

Moriya, S., Nakahira, K., Harada, M., Shino, M. and Kitajima, M.
Can Pupillary Responses while Listening to Short Sentences Containing Emotion Induction Words Explain the Effects on Sentence Memory?.
DOI: 10.5220/0011698200003417
In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) - Volume 2: HUCAPP*, pages 213-220
ISBN: 978-989-758-634-7; ISSN: 2184-4321

This paper focuses on narration, a component of content, to establish a content design method for more effective lateral knowledge transfer. When a narration comprises very short sentences and includes some elements that induce emotion, we assume that emotion is generated when a person hears it. Based on the relationships between (1) pupillary response and emotion, and (2) emotion and memory, which have been studied in recent years, we analyzed the possibility of capturing the characteristics of individual learners' acquisition of lateral knowledge through pupillary response, which is a measurable quantity. The results could contribute to the development of a content design method that is useful for transferring lateral knowledge. Section 2 describes the model used to conductthis study. Section 3 describes the experimental system based on the model. Section 4 presents the experimental results. In Section 5, based on the experimental results, we discuss the relationship between emotion, memory, and and pupillary response.

2 MODEL

2.1 Emotion and Memory

Murakami et. al. shows the cognitive model for memory formation introduced previous research (Murakami et al., 2021). Perceived information that has passed through the sensory memory and sensory information filter reaches long-term memory. Various types of chunks are stored in the long-term memory, and the chunks related to the input perceptual information are integrated in the working memory to produce reactions and make decisionsEach chunk is then enhanced by associating it with the perceptual information at the time the chunk was invoked.

One candidate for association would be the emotion that was generated when the chunk was invoked. The emotion that is generated may be a factor to improve memory performance. In general, it is difficult to control one's emotions. However, it is possible to add words that easily induce emotions to sentences, or to make it easier to induce emotions from the atmosphere of the sentences or induce emotions from the atmosphere of the sentences as a way to provide emotional stimuli. Based on this, we hypothesized that it is possible to investigate the relationship between the ease of remembering given information and emotion based on the biological responses of people when they perceive auditory stimuli that include emotion induction words.

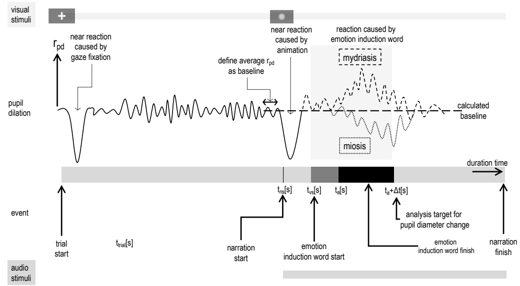

Figure 1: Behavioral model for pupil response.

2.2 Emotional Stimuli and Pupillary Response

The relationship between the emotion induction words as the auditory stimuli and the physiological responses to hearing them, is as follows: First, the emotion induction word is that which is related to emotion by its meaning. Regarding emotion, it is proposed that it can be expressed in two axes, pleasure/misery and arousal/sleepiness (Russell, 1980). Emotions such as emotion induction words are measured by a self-assessment manikin (SAM) (Bradley and Lang, 1994). This suggests that emotion can be represented on two axes, valence and arousal.

When using stimuli with emotional information, Affective Norms for English Words (ANEW) are used (Bradley and Lang, 1999). For each ANEW, valence, arousal, and dominance are measured as quantities indicating the degree of emotion. In Japanese, valence and arousal are measured against Japanese translations of words appearing in the ANEW (Honma, 2014). The words in this list were defined as Emotion Induction Words (EIW). Both valence and arousal are psychometric quantities as they are measured by the SAM. A person's psychological state is thought to affect his or her physiological state, including the pupil. Therefore, we considered that a person's physiological state is affected by the valence and arousal values of the emotion induction word, and we considered it appropriate to derive a relationship between the pupillary response and the valence/arousal of the EIW, which has been established previously.

In this paper, we considered a person's pupillary response to auditory stimuli containing EIWs as follows. Many studies have investigated pupillary responses and visual/auditory stimuli, emotion, and arousal (Zekveld et al., 2018). The pupillary response during word processing becomes larger when difficult or unknown words or words that convey darkness are read. We consider the relationship between EIWs and pupillary response in the following framework.

2.3 Quantifying Pupillary Responses to Emotional Stimuli

Figure 1 shows the model of stimulus - event timeline - pupil dilation per trial. The event starts at t_{trial}, but a sufficient time interval, t_{ns}, between the event and the time when the auditory stimulus narration starts so that the near reaction in the pupil does not interfere is required with the analysis of the pupil's response. The narration contains an EIW, and t_{vs} is made to utter the EIW after a certain time has elapsed from t_{ns}. When a person perceives auditory EIW, a biological reaction is induced according to the meaning, valence, and arousal. The generation of a response requires some time after the EIW is perceived. In this paper, we assume that this occurs around t_a and focus on the amount of change in pupil diameter $r_{pd}(t)$ at t among the pupillary responses that occur between t_a and Δt.

The pupil diameter change is calculated by the following formula with baseline $\tilde{r}_{pd}$ in the range ($t_{ns} - \Delta$[ms], t_{ns}) :

$$\tilde{r}_{pd} = \frac{1}{\Delta} \int_{t_{ns}-\Delta}^{t_{ns}} r_{pd}(t)\ dt$$

In this study, Δ was set to 500 ms. In this case, the pupil diameter change Δr_{pd} can be expressed by the following formula:

$$\Delta r_{pd}(t) = r_{pd}(t) - \tilde{r}_{pd}$$

The pupil diameter variation with Δr_{pd} from time t to δt can be expressed as:

$$\delta r(t) = \Delta r_{pd}(t + \delta t) - \Delta r_{pd}(t)$$

A $\delta r(t) > 0$ indicated mydriasis and $\delta r(t) < 0$ indicated miosis. The total change in mydriasis (r_{myd}) and total change in miosis (r_{mio}) can be expressed by the following equations:

$$r_{myd} \ \text{or}\ r_{mio} = \int_{t_a}^{t_a+\Delta t} \delta r(t) dt$$

The total change in pupil diameter, r_{all}, can be calculated using the following formula :

$$r_{all} = \text{abs}(r_{myd}) + \text{abs}(r_{mio})$$

We analyze the relationship between the subjective evaluation of sentences containing EIWs (r_{myd}, r_{mio}, r_{all}, and auditory stimuli) and the memory of the heard sentences and obtain knowledge about the design of easily remembered sentences.

3 EXPERIMENT

In the experiment, short sentences containing EIWs were presented as auditory information, and participants were required to evaluate their subjective impressions. After all sentences were presented, participants were required to perform a recall test for auditory information after a short break. In this series of procedures, the experimental environment was constrained to avoid affecting the participant's subjective impression of auditory information and the change in pupil diameter. This is further explained in Section 3.1. To more accurately estimate the relationship between the two features of the EIW (valence and arousal) and pupil response, it is important to design short sentences that include auditory stimuli. The method is explained in Section 3.2. The experimental procedure is explained in Section 3.3.

3.1 Method and Participants

The experiment was conducted in a soundproofed room with relatively low ambient noise. As the lighting condition in this room was maintained, the changes in participants' pupil diameter were not affected by ambient light. Participants received auditory stimuli through headphones connected to a computer. A Dell S2440L 24-inch display was used to project the experimental instruction and tasks. The participants were seated 0.8 m from the display and ask to stare at it during the experiment. Twenty-one participants in their 20s were included. Pupil diameter was measured using Tobii Pro Nano, a noninvasive corneal reflectometry system.

3.2 Design of Short Sentences

Auditory stimulus was designed based on the following four considerations: (1) whether it is comprehensible after one hearing, (2) classification of EIWs, (3) control of t_{ns}, t_{vs}, t_a, $t_a + \Delta t$ in Figure 1, and (4) control over the mood of the sentence.

Consideration (1) is important from the perspective of semantic comprehension and memory. Consideration (2) is important from the perspective of explaining the relationship between pupillary response and EIWs, which have two variables, valence and arousal, to connect observable value of bioinformatics. Consideration (3) is important to show that the pupillary response is definitely caused by the EIW. Consideration (4) is important from for clarifying the effect on subjective evaluation whether or not depending on the sentences which have different atmosphere but include same EIW.

First, the policy for designing short sentences to facilitate text comprehension by participants were as follows One consideration for facilitating auditory text comprehension is the length of each auditory

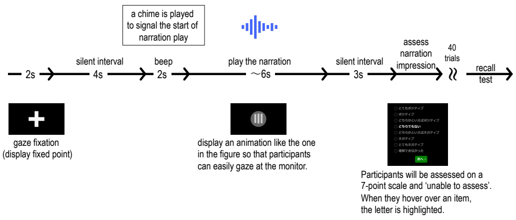

Figure 2: The experimental Design.

stimulus. Preliminary experiments showed that the length of auditory stimulus that participants could easily understand without any resistance was around 6 seconds. Therefore, we set the length of the sentence, an auditory stimulus, to around 6 seconds. This implies that an auditory stimulus must include approximately 30 ± 10 of kana characters, which were converted to mora numbers (counted in kana characters for Japanese). Moreover, most of professional announcers in Japan are trained to do so. Based on these facts, the number of kana characters (number of mora) around 6 seconds auditory stimulus should be included 1/10 of 300 mora.

The classification of EIWs is as follows. According to Honma's study (Honma, 2014), both valence and arousal for words that can be EIWs are classified in 9 levels. The set of EIWs is $\boldsymbol{W}$, where N_{EIW} represents the number of elements in $\boldsymbol{W}$:

$$W = \{W_i \mid i = 1, \cdots, N_{EIW}\}.$$

W_i has the average and standard deviation of the valence V_i and arousal Ar_i. Then, W_i is classified according to the following levels using valence V_i and arousal Ar_i.

V_i was classified as follows:

- Positive valence (V_+) 7.00 ~ 9.00
- Neutral valence (V_N) 4.00 ~ 6.99
- Negative valence (V_-) 1.00 ~ 3.99

V_+, V_N, V_- were further classified as follows: V_{++} $(7.70 \sim 9.00)$, V_{NN} $(5.00 \sim 5.99)$, V_{--} $(1.00 \sim 1.99)$.

Similarly, Ar_i was classified as follows:

- High arousal (Ar_H) 7.00 ~ 9.00
- Medium arousal (Ar_M) 4.00 ~ 6.99
- Low arousal (Ar_L) 1.00 ~ 3.99

Ar_H and Ar_L were adopted. For Ar_H or Ar_L, we used words with a standard deviation of 2.00 or less to suppress the variation in participants' impression of the stimuli when there are multiple words that satisfy the criteria. In addition, only one EIW was included in one sentence.

Next, the design policy for controlling t_{vs}, $t_{vs} + \Delta t$ as the timing of EIW appearance was as follows.

Table 1: Overview of designed auditory stimuli.

valence level of emotion-induction word: *POSITIVE* $(++)$

arousal level of emotion-induction word	atmosphere of sentence	number of sentences	notation for the sentence
HIGH	*POSITIVE*	3	(V_{++}, Ar_H, At_+)
HIGH	*NEGATIVE*	3	(V_{++}, Ar_H, At_-)
LOW	*POSITIVE*	3	(V_{++}, Ar_L, At_+)
LOW	*NEGATIVE*	3	(V_{++}, Ar_L, At_-)
total number of sentences		12	

valence level of emotion-induction word: *NEUTRAL* (N)

arousal level of emotion-induction word	atmosphere of sentence	number of sentences	notation for the sentence
HIGH	*POSITIVE*	2	(V_N, Ar_H, At_+)
HIGH	*NEUTRAL*	4	(V_N, Ar_H, At_N)
HIGH	*NEGATIVE*	2	(V_N, Ar_H, At_-)
LOW	*POSITIVE*	2	(V_N, Ar_L, At_+)
LOW	*NEUTRAL*	4	(V_N, Ar_L, At_N)
LOW	*NEGATIVE*	2	(V_N, Ar_L, At_-)
total number of sentences		16	

valence level of emotion-induction word: *NEGATIVE* $(--)$

arousal level of emotion-induction word	atmosphere of sentence	number of sentences	notation for the sentence
HIGH	*POSITIVE*	3	(V_{--}, Ar_H, At_+)
HIGH	*NEGATIVE*	3	(V_{--}, Ar_H, At_-)
LOW	*POSITIVE*	3	(V_{--}, Ar_L, At_+)
LOW	*NEGATIVE*	3	(V_{--}, Ar_L, At_-)
total number of sentences		12	

We set $t_{ns} \sim t_{vs}$ to an interval of ~ 1 second in order to easily capture the pupillary response. Therefore, around 5 mora auditory such as word and so on set before EIW. To prevent Δt from being too short, the number of mora in EIW was also set to be around 5 mora. In addition, $t_a - t_{vs}$ were set at 1 [s].

Finally, the design policies for the atmospheres of the sentences were as follows. The atmospheres of the sentences were classified as *Positive*, *Neutral*, and *Negative*, and represented by At_+, At_N, and At_- respectively. The atmospheres of the whole sentences were generated based on the following policies.

- For At_+, EIWs with V_{++} were used, and positive vocabulary or double negation were used at the end of sentences. When an EIW with V_{--} was used, the word was not used as the subject but to modify other vocabulary.
- For At_N, objective facts were described without using emotive words.
- For At_-, negative vocabulary was used at the end of the sentence in addition to EIW; positive vocabulary was not used in the sentence or positive vocabulary was used in the sentence and negated at the end.

Based on the above, we generated 40 auditory stimuli with the combinations shown in Table 1.

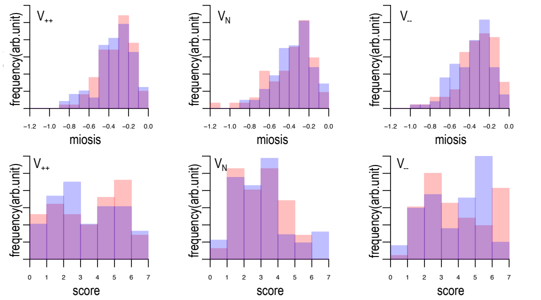

Figure 3: Histograms of miosis distribution (upper row) and the impression evaluation value (lower row) for all participants categorized by valence and arousal. From left side to right side, it is for V_{++} / V_N / V_{--}. Red histogram is for high arousal, and blue histogram is for low arousal.

Table 2: Categorization of participants' types which combinations are $V_{++}/V_N/V_{--}$, Ar_H / Ar_L, $Ar_+/Ar_N/Ar_-$.

	V_{++}, Ar_H			V_N, Ar_H			V_{--}, Ar_H		
	assess	recall	closed	assess	recall	closed	assess	recall	closed
At_+	47(74.6)	0	17(27.0)	40(95.2)	0	24(57.1)	46(73.0)	8(62.5)	28(44.4)
At_N	N/A	N/A	N/A	36(42.9)	1(100.0)	43(51.2)	N/A	N/A	N/A
At_-	53(84.1)	0	35(55.6)	30(71.4)	0	23(54.8)	40(63.5)	7(71.4)	25(39.7)
	V_{++}, Ar_L			V_N, Ar_L			V_{--}, Ar_L		
	assess	recall	closed	assess	recall	closed	assess	recall	closed
At_+	54(86.0)	0	32(50.8)	31(74.0)	0	14(33.3)	43(68.3)	1(0.0)	33(52.4)
At_N	N/A	N/A	N/A	28(84.0)	2(0.0)	42(50.0)	N/A	N/A	N/A
At_-	49(77.8)	5(100)	34(54)	32(76.2)	1(0.0)	21(50.0)	50(79.4)	0	31(49.2)

3.3 Procedure

The experimental flow of each trial is described based on Figure 2. First, gaze fixation was performed without auditory stimuli for 2 [s] at the beginning of the experiment to fix the pupil diameter measurement. Then, the soundless interval was set to 4 [s] to avoid any influence on the pupil diameter measurement between the presented stimuli, assuming a near-sighted pupil effect.

Next, a chime sound was presented for 1 [s] to indicate the start of the experiment, followed by a 1 [s] soundless interval to prevent any change in pupil diameter caused by the biological response to the chime sound. Later, the narration, which is the presented stimulus of the experiment, was replayed. The reason for presenting animation at the same time was to make it easier for the participant to fixate his/her gaze. We confirmed via preliminary experiments that animation has little effect on the mydriasis or miosis of the pupil diameter. At this time, the animation shown in Figure 2 continued to be displayed. A soundless interval of 3 seconds is set after the narration was complete to stabilize the pupil diameter fluctuation, so that the pupil diameter measurement during the narration impression evaluation would not be affected.

Finally, the subjective impression of the narration on the participants was conducted according to a 7-points Likert scale shown on the computer screen. The subjective evaluation was performed by selecting the rating results using radio buttons, as shown in Figure 2. When the selection was complete, the "Next" button was clicked to start the next trial. All visual and auditory stimuli before the subjective impression evaluation screen were presented automatically. After repeating the above 40 trials, a recall test was conducted by asking the participant to recite as many of the 40 sentences as possible from memory.

4 RESULT

The experimental results are summarized in Table 2 and Figure 3. Table 2 shows the responses of participants classified by $(V_{++} \vee V_N \vee V_{--})$ and $(Ar_H \vee Ar_L)$, which are characteristics of EIWs included in sentence stimuli and $(At+/At_N/At_-)$, which are the characteristics of the whole sentence. The column "assess" shows the subjective impressions of the sentence stimuli, and are answered in the same range as $At_+/At_N/At_-$. The value indicates the number of respondents, and the numbers in parentheses indicate the percentage of all trials. The column "recall" shows the results of the recall test. The value in the "recall" column indicates the number of people who remembered the corresponding characteristic. The values in parentheses indicate the percentage of those who remembered the content of the short sentences whose subjective impression rating matched the characteristic assigned to the sentence stimuli. The column "closed" shows the results of those whose pupillary response tended toward miosis. The value in the "closed" column indicates the number of subjects whose pupillary response showed miosis tendency, and the number in parentheses indicates the percentage of the total trials.

Figure 3 shows the frequency distribution of the pupil response and subjective evaluation of the stimuli. These were generated using data from 658 trials, excluding data for pupillary response with a pupil diameter acquisition rate of less than 70% per trial and an outlier rate of 30% or more. The histograms for miosis are shown in the upper row of Figure 3, and the histograms for impression assessment are shown in the lower row. The columns of the graph are divided by the EIW characteristics included in the sentence stimuli and are plotted against $V_{++} \vee V_N \vee V_{--}$ from left to right. The histograms in red and blue in the figure represent Ar_H and Ar_L, respectively.

Table 3: p-values of r_{mio}, r_{myd}, and r_{all}. 0.05, 0.1, and effective in column name represents categories for sentence sampling. The parameter combination represents left and right side of comma.

	miosis			mydriasis	total change	
p-value	0.05	0.1	effective	0.05	0.05	effective
V_{++}, V_{--}	0.0030	-	0.0189	-	0.0526	-
V_{++}, V_N	0.0057	-	-	0.0282	-	-
V_{--}, V_N	-	0.097	-	-	-	-
$V_{++}Ar_H, V_NAr_H$	0.0038	-	0.0044	-	-	-
$V_{++}Ar_L, V_NAr_H$	0.0315	-	0.0626	0.0369	-	-
$V_{++}Ar_L, V_NAr_L$	-	-	-	0.0792	0.0173	-
$V_{--}Ar_H, V_NAr_H$	0.0240	-	0.0055	0.0553	-	-
$V_{--}Ar_H, V_NAr_L$	0.0201	-	0.0105	-	-	-
$V_{--}Ar_L, V_NAr_L$	-	-	-	-	-	0.0565
V_NAr_H, V_NAr_L	-	-	0.0937	-	0.0068	-

4.1 Preliminary Analysis

First, we explain the results shown in Table 2. The recall test was conducted orally, and the participants were asked to report what they remembered in the short sentences presented as stimuli, using as many words and phrases as possible. If the reported content approximated the presented sentence, the sentence was judged to have been memorized. Six participants reported one sentence, 10 reported two, two reported four, and none reported three sentences.

The relationship between V_i, Ar_i, At_j and recall was as follows. The recall rate was highest for short sentences containing EIWs characterized by $V_{--} \wedge Ar_H$, regardless of the atmosphere of the short sentence. Short sentences characterized by $V_{++} \wedge Ar_L \wedge At_-$ also had reasonably high recall rates. These findings support the results of a previous study (Murakami et al., 2021).

The relationship between pupillary response and recall was as follows. The relationship between the valence of EIWs and the mydriasis rate was $40 \sim 54\%$ for $V_{++} \vee V_{--}$ and $28 \sim 40\%$ for V_N. This result indicates that mydriasis is more likely to occur in case of $V_{++} \vee V_{--}$. In terms of miosis, V_{++} in Table 2 shows a strong trend toward miosis in $V_{++} \wedge Ar_H \wedge At_+$. Of the 15 participants who were able to recall short sentences with EIWs of $V_{--} \wedge Ar_H$ in the recall test, a total of 7 participants had a pupillary response of a trend toward miosis.

Next, we explain the results shown by Figure 3. Looking at the density distribution for miosis (upper row of the figure), the values ranged from -0.3 to -0.2, However, there is a slight difference in the distribution of pupil response between Ar_H and Ar_L. Especially for V_{++}, the overall miosis value was larger than -0.5. But for V_{--}, compared to Ar_H or Ar_L shows a slight advantage in the area where it takes values less than -0.5. Furthermore, the density distribution of subjective impression assessment for short sentences (lower row of the figure) shows a large difference between $Ar_H \wedge At_-$ and $Ar_L \wedge At_-$, especially

Table 4: Average of r_{mio}, r_{myd}, and r_{all}. 0.05, 0.1, and effective in column name represents categories for sentence sampling. Parameter representation is same as Table 3.

pupil response	miosis			mydriasis	total change	
	0.05	0.1	effective	0.05	0.05	effective
V_{++}	-0.2869	-	-0.3236	0.4012	-	-
V_N	-0.3937	-0.3673	-0.3773	0.3302	0.6826	-
V_{--}	-0.3932	-0.3176	-	-	0.7768	-
$V_{++}Ar_H$	-0.279	-	-0.311	-	-	-
$V_{++}Ar_L$	-0.305	-	-0.340	0.413	0.762	-
$V_{--}Ar_H$	-0.298	-	-0.310	0.400	-	-
$V_{--}Ar_L$	-0.422	-	-0.400	-	0.777	0.734
V_NAr_H	-0.438	-	-0.412	0.295	0.748	-
V_NAr_L	-	-	-0.354	0.341	0.608	0.658

for V_{--}. Compared to Ar_L, the subjective impression assessment of Ar_H is split into positive/negative, whereas the negative impression of Ar_H is expected to be negative. The evaluation of sentences with EIW for V_N generally shows a positive value.

4.2 Valence–Arousal of Sentences and Pupil Diameter Change

We computed r_{mio}, r_{myd}, and r_{all} for each participant and each sentence using the method described in Section 2 and analyzed their behavior toward emotional stimuli. The sensitivity of the sentences to r_{mio}, r_{myd}, and r_{all} is unknown. We performed a t-test across all 40 sentences for each pupillary response, and analyzed the emotional stimuli with $p < 0.1$. The number of sentences sampled for r_{mio}, r_{myd}, and r_{all} were $N_{s,mio} = 24$, $N_{s,myd} = 27$, and $N_{s,all} = 24$, respectively. We represent the sampling sentence groups by $G_{s,mio}$, $G_{s,myd}$, and $G_{s,all}$, respectively. Each group comprised sentences extracted at three levels: $p < 0.05$, $0.05 < p < 0.1$, $p < 0.1$. We represent the attribute of each group by $G_{s,mio/myd/all,p-value}$. For example, $G_{s,mio,0.05}$ was analyzed for r_{mio} using the group of 24 sentences in $N_{s,mio}$ that satisfied $p < 0.05$.

Each emotional stimulus contained EIWs with specific V_i, Ar_i. Therefore, if r_{mio}, r_{myd}, and r_{all} have some relationship with EIWs, they should have some characteristics for valence or arousal. Therefore, we extracted r_{mio} for $G_{s,mio}$, r_{myd} for $G_{s,myd}$, and r_{all} for $G_{s,all}$, and calculated the $(V_{++}/V_N/V_{--})$. (Ar_H/Ar_L) of the sentences were combined to perform the t-test.

Table 3 shows the results of the tests for the combined EIW characteristics that were significantly different or tended to be significant. Several trends were observed among the valences. For r_{mio}, significant differences or significant trends were found for all combinations of V_{++}/V_{--}, V_{++}/V_N, and V_{--}/V_N. On the other hand, r_{myd} showed significant differences only among V_{++}/V_N, while r_{all} showed no significant differences or trends. This indicates that miosis or mydriasis is suitable for valence state identification. No significant difference or trend was found for

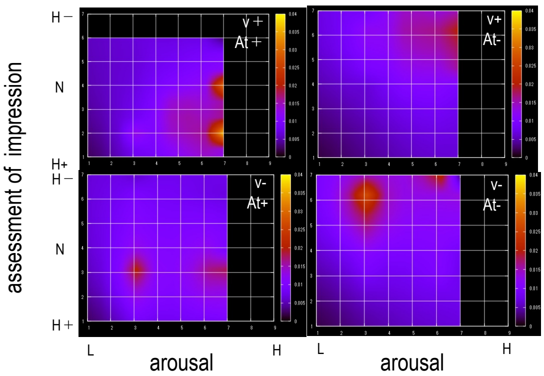

Figure 4: Valence, arousal, atmosphere, assessment of impression. V_+ / V_- represent high/low valence, At_+ / At_- represent positive/negative atmosphere.

arousal. The same is true for the results of the t-test for the valence/arousal combination in the table.

Table 4 shows the average of $r_{mio}/r_{myd}/r_{all}$ for the conditions with significant differences or significant trends in Table 3. The trend for miosis against valence was generally negative with a large V_N r_{mio} and a small V_{++}/V_N. This indicates that the response of miosis is weak for V_{++}/V_{--}. On the other hand, V_{++} shows a larger positive value for mydriasis. The trend of miosis for the valence/arousal combination $V_{++} \wedge Ar_H$, $V_{--} \wedge Ar_H$ shows small negative values. This indicates that the response of miosis is weak for $Ar_H \wedge$ $(V_{++} \vee V_{--})$. For $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$, mydriasis showed large positive values. This indicates that the response of mydriasis is strong for $V_{++} \wedge Ar_H$ or $V_{--} \wedge Ar_L$. In summary, the common trend was miosis/mydriasis response to $V_{--} \wedge Ar_H$.

5 DISCUSSION

As mentioned in section 2, many studies have indicated a certain relationship between pupillary response and emotion. Lavoie and O'Connor suggest that, for memory and emotion, emotional events are associated with a better recall than that associated with non-emotional events, and that ERPs reflecting temporal conscious recall and post-retrieval monitoring are clearly affected by both valence and arousal (Lavoie and O'Connor, 2013). Gomes et al. used a dual process model to investigate the effects of valence and arousal on memory, suggesting that valence and arousal affect memory in relation to each other but not in isolation (Gomes et al., 2012). Megalakaki et al. conducted a remember/known test on text comprehension and memorization in which valence and emotional intensity were individually manipulated. The results suggest that emotion plays an important role in memorization as positive/negative content is more easily recalled than neutral content and that positive words are more easily recalled (Megalakaki et al., 2019).

Based on the results of the experiment, the pupillary response and its effect on sentence memory when a short auditory stimulus containing EIW is perceived may be attributable to the following:

- Relationship between valence, arousal, pupillary response, and memory, i.e., the features of EIW
- Relationship between the EIW feature values, atmospherics, and subjective impression assessment of the auditory stimulus.

5.1 Relation Between Memory and Pupil Response for EIWs

The results presented in Table 3 suggest that $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$ can be separated from V_N. Table 4 shows that $\bar{r}_{mio}$ for $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$ is -0.279 and -0.298 respectively, showing a lower miosis tendency than did other states. The $\bar{r}_{mio}$ for $V_{++} \wedge Ar_L$ was -0.305, again showing a low miosis tendency. Conversely, $\bar{r}_{myd}$ for $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$ is 0.413 and 0.400, which is higher than that in other states. This shows a good correspondence with the high memory for $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$ sentences in Table 2. These results suggest that valence and arousal have a certain influence on human memory, and that sentences containing EIWs with emotions expressed by $V_{--} \wedge Ar_H$ and $V_{++} \wedge Ar_L$ are easily remembered. This also suggests that measuring the pupillary response of people who listen to auditory stimuli can be used to estimate the likelihood of emotion felt on perceiving auditory stimuli, and thus the degree to which they are likely to remember them.

5.2 EIW Features, Atmospheres, and Subjective Impression Assessment

This section discusses which valence, arousal, and atmosphere included in the auditory stimulus is most likely to influence the subjective impression evaluation. Figure 4 shows the heatmap of two-dimensional frequency distribution related to arousal derived from EIWs in short sentences and the score of the subjective assessment of the impression for short sentence as auditory stimuli. Each short sentences including EIWs characterized by V_i, Ar_i, and At_j. The horizontal axis of each graph represents arousal, which changes from low to high arousal from left to right. The vertical axis represents the assessment of impression, which changes from *positive* $\sim$ N $\sim$ *negative* from the bottom to the top of the axis. The graph comprehension is as follows: the lower right region of the heatmap shows Ar_H and positive assessment of impression, and the upper right region shows Ar_H and negative assessment of impression. As shown in Table 1, the number of auditory stimuli classified by $(V_i,\ Ar_i,\ At_j)$ is almost equal. Therefore, the anticipated distribution of subjective impression results in Figure 4 comprises approximately equivalent peaks for each feature of auditory stimulus. For example, $(V_{++},\ At_+)$ has two strong distributions around $(Ar_i,\ S_{ai}) \sim (Ar_L \vee Ar_H,\ 1 \sim 2)$ and is expected to have more strongly positive (here $S_{ai} \sim 1$) results for Ar_H than for Ar_L.

Overall, these results suggest that the short sentence impression expected from the combination of $(V_i,\ Ar_i,\ At_j)$ and the participant's subjective impression assessment are generally consistent for short sentences. However, short sentences with the $(V_{++},\ Ar_H,\ At_+)$ feature require more work. This indicates that, at least for EIW, it is sufficient to check V_i and Ar_i and incorporate them into short sentences. We will investigate the relationship between At_j and pupillary response to understand the relationship between subjective impression assessment of the short sentence and pupillary response. Further, this suggests that the relationship with memory can be referred to in the future.

6 CONCLUSION

The purpose of this study was to detect relationship between the response of individual participants to the EIW characterized by valence and arousal, and their memory of short sentences. We designed a short sentence containing an EIW as an auditory stimulus to facilitate measurement of the pupil dilation response, based on the idea that the EIW could be characterized by valence and arousal and related to emotion. Participants were then presented with auditory stimuli, and their impressions of the auditory stimuli, pupil dilation response, and remembered sentences were measured. The results suggest that the $r_{myd}\ /\ r_{mio}$ of auditory stimuli, such as narration, can be easily memorized by measuring the content of the auditory stimuli. Based on this finding, it will be possible in the future to create and present narration-like commentaries tailored to individual characteristics.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS KAKENHI Grant Number 19K12232, 19K12246, 20H04290, and 22K12284. MH also wants to thank to Nagai N · S Promotion Foundation For Science of Perception for their finantial support.

REFERENCES

Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59.

Bradley, M. M. and Lang, P. J. (1999). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.

Gomes, C., Brainerd, C., and Stein, L. (2012). Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of experimental psychology. Learning, memory, and cognition*, 39:663–677.

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 the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: HUCAPP*, pages 162–169. INSTICC, SciTePress.

Honma, Y. (2014). Drawing up of the japanese word stimulus based on the emotional valence and arousal of the word. *Bulletin of Aichi Institute of Technology*, 49:13–24.

Lavoie, M. and O'Connor, K. (2013). Effect of emotional valence on episodic memory stages as indexed by event-related potentials. *World Journal of Neuroscience*, 03:250–262.

Megalakaki, O., Ballenghein, U., and Baccino, T. (2019). Effects of valence and emotional intensity on the comprehension and memorization of texts. *Frontiers in Psychology*, 10.

Murakami, M., Shino, M., Nakahira, K., and Kitajima, M. (2021). Effects of emotion-induction words on memory of viewing visual stimuli with audio guide. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: HUCAPP*, pages 89–100. INSTICC, SciTePress.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: Current state of knowledge. *Trends in Hearing*, 22:2331216518777174. PMID: 30249172.