Local-Global Reaction Map for Classifying Listeners by Pupil Response to Sentences with Emotion Induction Words and Its Application to Auditory Information Design

Katsuko T. Nakahira*, Munenori Harada*, Shunsuke Moriya*, Muneo Kitajima*
* *Nagaoka University of Technology*
Nagaoka, Niigata, Japan
Email: {katsuko@vos, s193369@stn, s213345@stn, mkitajima@kjs}.nagaokaut.ac.jp

***Abstract*— When a person acquires a text as auditory information and derives the meaning of the text, he or she may simultaneously generate an emotion in response to the content of the text. Emotions are said to have a certain relationship with decision-making and memory. Therefore, it is expected that even sentences with the same meaning will be remembered differently depending on the emotion evoked. This study aims to clarify the relationship between the emotions that arise when listening to a text and the memory of the presented text. The classification of emotional states held by people is performed by a method based on subjective quantities by impression rating or by a method based on objective quantities by biometric information. In this study, we focus on pupil response, which is biological information that has been suggested to change with emotion. Based on this, this paper proposes the Local-Global Reaction Map (LGR-Map) as a classification method for pupil changes accompanying emotional changes, as a basic research for the construction of adaptive content design methods that utilize the degree of human emotional arousal. The LGR-Map is generated by capturing the emotional changes during listening to a text from the following two perspectives; Those generated by words in a specific region of a sentence (Local reaction); those generated by the context of the entire sentence (Global reaction). The total pupil diameter change within a certain time period is obtained as the characteristic quantity for each response. Error ellipses are defined for the distribution of listeners in the LR-GR for the presented text (LGR-Map), and classified into five types based on the rotation angle and flattening ratio of the error ellipses. The basic properties of the LGR-Map were investigated by using auditory stimuli presented in short sentences containing Affective Norm for English Words (ANEW). As an extension, we will attempt to create an extended LGR-Map for sentences with multiple ANEWs and consider whether it is possible to extract features of the pupillary response. In addition, we discuss the consistency of the results of a recall test in relation to the cognitive model.**

Keywords*— *Local-Global Reaction Map; Pupil Response; Affective Norm for English Words; Emotion Induction; Contents Design of Auditory Information.

I. INTRODUCTION

This paper is based on the previous work originally presented in COGNITIVE2023 [1]. An extended LGR-Map analysis was added in Section V.

With the penetration of mobile devices and the development of eXtended Reality (XR) technology, we are surrounded by an increasing number of services that disseminate content via electronic media. Many of these services are designed to enrich the experience of individuals, and their range of application is wide, from sensory experiences, such as sightseeing and movies to educational materials that make it easier for people to acquire knowledge. In recent years, there has been a movement to expand content provision services from an inclusivity perspective (e.g., Vallez et al. [2]).

Content design is essential to content provision in the sense of striving to convey what is to be conveyed as accurately as possible. Content design has the issue of the quality and quantity of the presenting stimulus as the material contained in the content. Visual and auditory information are the central presenting stimuli, and how to handle their quality and quantity is one of the key factors.

Regarding the amount of content, since perceptual information is basically a physical quantity, the amount of processing is determined by the structure of the human cognitive system itself, and individual differences are usually negligible. This is described by Hirabayashi et al. [3] as the relationship between the amount of information and the timing at which the information is given, and it is possible to maximize human memory by giving visual and auditory information, or explicit and implicit information in the appropriate order and intervals.

Furthermore, the quality of content is largely related to the viewer's cognitive process. The cognitive process depends on the richness of information nodes and the state of node connectivity of the information receiver, and thus varies from person to person. Murakami et al. [4][5] discussed the quality of content for short auditory information. They classified the emotions of short sentences into positive, negative, and other categories (in this case, we assign neutral), and calculated memory scores for each category, suggesting that short sentences belonging to a specific category improve memory scores. They also suggested the possibility of using pupil response to measure human emotion induction from short sentences. In addition, Moriya et al. [6] found that pupil responses to Affective Norm for English Words (ANEW) contained in short auditory information may be characterized based on ANEW categories.

Therefore, in order to design content that facilitates better emotional experiences and knowledge acquisition, it is desirable to be able to adaptively provide content according to the viewer's cognitive characteristics. For this purpose, it is necessary to monitor the viewer's emotional state in real time. Biometric information is a suitable indicator for this purpose. There are many types of biometric information on emotion (e.g., Jim et al. [7], Shu et al. [8]), but considering the time scale and ease of measurement, the pupillary response is the most promising. Based on the above, this paper focuses on

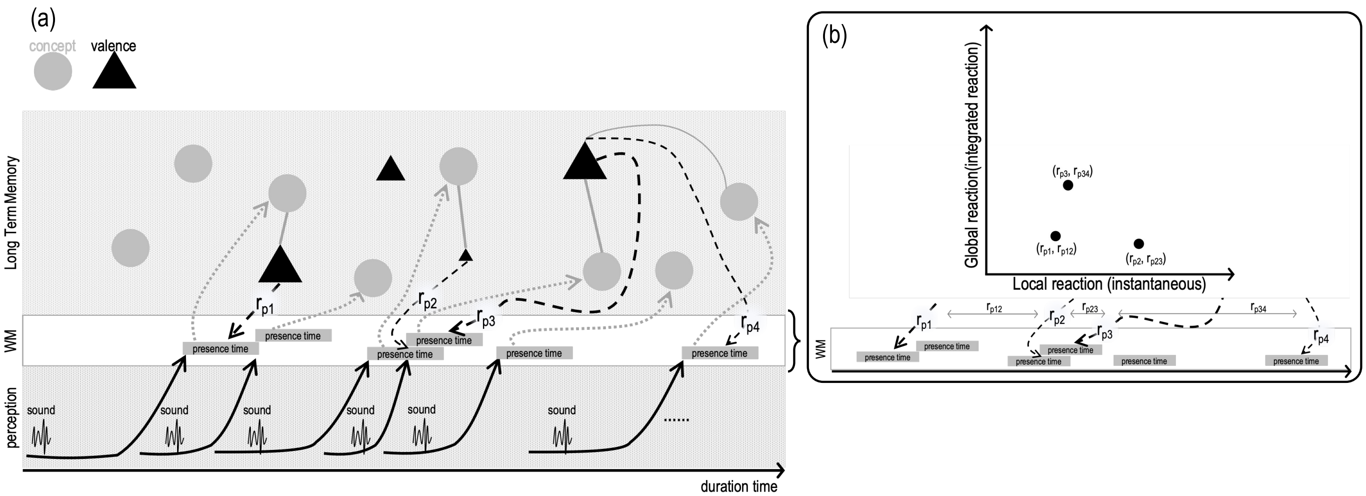

Figure 1. Cognitive model of this paper based on CI model. (a) Input - Cognitive process. (b) Working memory processing - output process.

pupillary response and proposes the Local-Global Reaction Map (LGR-Map) as a classification method for pupillary changes associated with emotional changes.

This paper is organized as follows. In Section II, we construct a base cognitive model and propose an LGR-Map based on it. In Section III, we show the usefulness of the LGR-Map by actually applying it to the HUCAPP 2023 data [6]. In Section IV, we discuss the usefulness of LGR-Map. In Section V, as an extension of the LGR-Map, we attempt to interpret the relationship between pupillary response, and the result of recall test when multiple ANEWs are present.

II. Designing LGR-Map based on Construction-Integration Model

A. Basic Design

In this paper, we construct a reaction model for human emotion based on the Construction-Integration Model (CI-model, e.g., [9][10][11]) proposed by Kintsch. The CI-model is a theory of discourse comprehension consisting of a construction step and an integration step.

The scenario in this paper is modeled based on the CI-model as shown in Figure 1. Figure 1 (a) shows the construction process that encodes information (packet of sound waves) input from the outside world, retrieves information stored in long-term memory using it as a clue, and constructs a network. Figure 1 (b) shows the integration process in which the retrieved information is pruned and integrated in the working memory by pruning information that does not fit the context, and the physical response is output.

First, when a single stimulus (a packet of sound waves in the auditory case) is perceived from a sensory organ, it is sent as encoded perceptual information from the sensory organ to the working memory. The sent information is matched with a large number of nodes (knowledge concepts) in the brain's long-term memory. The corresponding knowledge concept and its associated knowledge concept are then returned to the working memory. In this case, the information of the chunk of emotion (defined by valence and arousal) r_{p_i} associated with the knowledge concept is also returned, so that the working memory temporarily retains the emotion of the perceived packet of sound waves. Based on the returned r_{p_i}, the cognitive process via the working memory activates the motor process in each part of the body, and a response is generated. The pupillary response we focus on in this paper is produced by the activity of the pupillary sphincter and pupillary dilator muscles, which are considered to be one of their responses. The story so far can be expressed as follows.

Let $\boldsymbol{K}$ be a row vector of $^{\forall}$word concepts (knowledge concepts) in the long-term Memory (LTM) of $^{\exists}$person, and the word concept i input at time t_j is denoted by the element $K_i(t_j)$ in $\boldsymbol{K}$. where $K_i(t_j)$ are the values of valence V_i and arousal Ar_i that characterize the emotion [12]. The number of elements is $i = 1 \cdots n_K$ (n_K is the total number of word concepts), with only one i value of 1 for some time t_j. Here, V_i or Ar_i or both may have no value (~ 0) (in that case, $V_i = 0,\ Ar_i = 0$). The range of values for V_i and Ar_i is $1 \leq V_i \leq 9$, $1 \leq Ar_i \leq 9$.

Next, let $\boldsymbol{A}$ be a column vector of $^{\forall}$emotion concepts in the LTM of $^{\exists}$people, consisting of elements $A_k(V_k,\ Ar_k)$. The number of elements is $k = 1, \cdots, m_A$ (m_A is the total number of emotion concepts), and there always exist $V_k,\ Ar_k$ values.

The $\boldsymbol{K}$ and $\boldsymbol{A}$ are connected by a $n \times m$ matrix $\boldsymbol{W(t_j)}$ that shows their connectivity at time t_j. The element $w_{ik}(t_j)$ of $\boldsymbol{W(t_j)}$ indicates the degree of coupling between $K_i(t_j)$ and A_k. If $K_i(t)$ in $\boldsymbol{K}$ is input and co-occurs with A_k in $\boldsymbol{A}$ on t_j, the probability that $K_i(t_j)$ retrieved from LTM is $p(K_i(t_j))$ and that A_k retrieved from LTM is $p(A_k(t_j))$, the probability of A_k being retrieved from LTM is expressed by the following equation.

$$w_{ik}(t_j) = w_{ik}(p(K_i(t_j)),\ p(A_k(t_j)))$$

In this case, the temporary emotion $E(t_j)$ generated from the

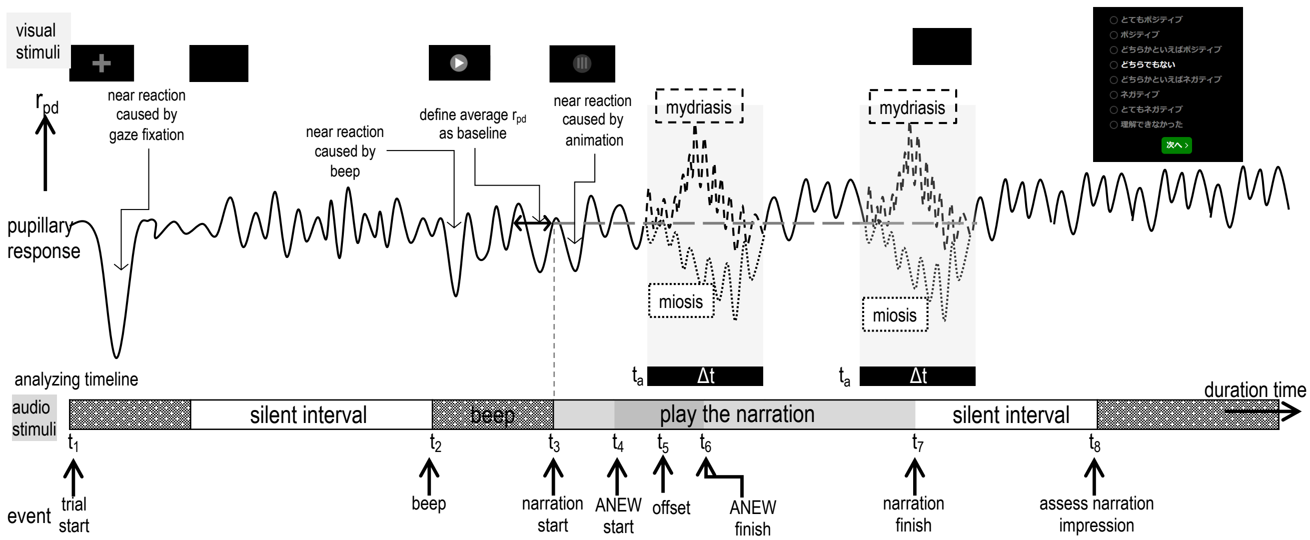

Figure 2. The full pupillary response when auditory information is given.

input $^\exists$packet of sound waves is (1).

$$E(t_j) = D(t_j) \sum_{i=1}^{n_K} \sum_{k=1}^{m_A} K_i(t_j) w_{ik}(t_j) A_k \tag{1}$$

Here, $D(t_j)$ is the damping factor. The above explanation represents the construction process.

The sentence, which consists of n_{wp} packets of sound waves, repeats the process of (1) as one cycle up to this point, and continues to return the emotion associated with the knowledge concept to the working memory. In the process, $E(t_j)$ may or may not be integrated between t_j depending on the presence or absence of active sources and contextual relations. The damping factor is introduced as a quantity that indicates the degree of such emotional integration. When n_{wp} packets of sound waves are listened to, the emotion arises in the form of integration of $E(t_j)$ that has been cultivated up to that point. It is usually at the end of a sentence where the packets of sound waves are interrupted. This is the integration process in this research situation.

Based on this, we consider the 2D plane shown in Figure 1(b). We thought that we could show the characteristics of the emotion that occurs in listeners when they listen to narration by plotting the information on human reactions in this plane. The r_{p_i} in the figure indicates the emotional reaction to a specific packet of sound waves. $r_{p_{i,i+1}}$ indicates the emotional reaction generated by the integration of the emotional reactions generated by multiple packets of sound waves. By treating it in this way, two measured reaction quantities can be plotted on a plane as $(r_{p_i},\ r_{p_{i,i+1}})$. In this paper, we call this plane as *Local-Global Response Map* (LGR-Map).

B. Representation of Pupillary Response based on CI-model

In order to apply the LGR-Map to pupillary responses, the measurement design of pupillary response should keep an adequate time interval by both inducing time interval both inducing a specific emotion induction word in narrating (instantaneous response) and context of narration (integrated response). Figure 2 represents a measurement design of pupillary response when listening to the narration stimuli based on Figure 1. Here, we adopt the Japanese version of ANEW [15], which induce emotion as the result of instantaneous response. For the integrated response, we assumed that the effect appears at the end of the sentence. We measured participants' pupillary responses to short sentences containing one ANEW word.

The auditory stimuli are adjusted for the event specified in Figure 2 as follows. The beep sound for the mental preparation to initiate auditory stimuli is uttered at t_2. Narration starts at t_3 and ANEW is uttered at t_4. After that, auditory stimuli are terminated at t_7. During this period, the auditory elements related to the evocation of emotion are the ANEW and the atmospheres at the end of the sentence. When analyzing the pupillary response, it is necessary to analyze the data in the vicinity of these elements.

Next, pupil diameter $r_{pd}(t)$ at elapsed time t is processed as follows, in the following order: determination of baseline, calculation of pupil diameter change, and total pupil diameter change.

First, when we set Δt_b as the interval necessary to calculate baseline, baseline $\tilde{r}_{pd}$ is calculated as follows.

$$\tilde{r}_{pd} = \frac{1}{\Delta t_b} \int_{t_{ns}-\Delta t_b}^{t_{ns}} r_{pd}(t)\ dt \tag{2}$$

Here, pupil diameter change value in t $\Delta r_{pd}(t)$ is calculated by the equation (3).

$$\Delta r_{pd}(t) = r_{pd}(t) - \tilde{r}_{pd} \tag{3}$$

The pupil diameter change $\delta r(t)$ between the duration time t and δt is calculated by the equation (4).

$$\delta r(t) = \Delta r_{pd}(t + \delta t) - \Delta r_{pd}(t) \tag{4}$$

Mydriasis (dilation) and miosis (constriction) are typical quantities that show pupillary response. Since the instantaneous changes in either of them are minute, we represent the total amount of change in only mydriasis or only miosis at $[t_a,\ t_a + \Delta t]$. These can be expressed as total amount of mydriasis r_{myd}. The total amount of miosis r_{mio} is calculated by the equation (5).

$$r_{myd} \text{ or } r_{mio} = \int_{t_a}^{t_a+\Delta t} \delta r(t)\ dt \tag{5}$$

In order to obtain a clearer picture of the change in pupillary response, it is better to capture the absolute change in $r_{myd},\ r_{mio}$. Here, we define r_{all} as the total change in pupillary response calculated by (6).

$$r_{all} = |r_{myd}| + |r_{mio}| \tag{6}$$

In LGR-Map, r_{all} is assumed to be a local (instantaneous) or global (integrated) reaction around calculated R. The pupillary response analysis start time t_a and analysis interval Δt can be arbitrarily determined. In the LGR-Map, we set t_a and Δt using the event time in Figure 2 as follows: For the local reaction, t_a is set to t_5, the offset time at which the pupil response is expected to start after the appearance of the ANEW that causes the reaction. For the global reaction, t_a is set where the integrated effect can be easily confirmed. In this paper, t_a is set a little before t_7, when the narration ends. Since the actual narration has n_s sets of calculated points or consists of n_s sentences, at most n_s points are plotted on the LGR-Map.

C. *Typology based on LGR-Map*

r_{all} distribution on LGR-Map is regarded as a description of induced emotion by narration stimuli for each participant. We design the method of categorization of typology for r_{all} distribution on LGR-Map. r_{all} distribution has x axis for local response and y axis for global response. r_{all} is the information including the individual differences. Now, each individual difference is assumed to obey a normal distribution. If the distribution obeys a two-dimensional Gaussian distribution, we can draw the error ellipsoid on LGR-Map.

The error ellipsoid is represented by the following equation using the transformed coordinates $u,\ v$. Hence $\sigma_u^2,\ \sigma_v^2$ are the variances of the transformed coordinates with respect to the respective axes.

$$\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} = c^2$$

Here, $\sigma_u^2,\ \sigma_v^2$, and rotation angle of error ellipsoid α can be converted as the equations (7) – (9) using σ_x^2 as variance for local response, σ_y^2 as variance for global response, σ_{xy} as covariance of local-global response.

$$\sigma_u^2 = \frac{\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}}{2} \tag{7}$$

$$\sigma_v^2 = \frac{\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4\sigma_{xy}^2}}{2} \tag{8}$$

$$\tan\alpha = \frac{\sigma_{xy}}{\sigma_u^2 - \sigma_y^2} \quad (0 < \alpha < 180^\circ) \tag{9}$$

We consider the shape of error ellipsoid depending on the behavior of $\sigma_x^2,\ \sigma_y^2,\ \sigma_{xy}$. First, we relate σ_x^2 and σ_y^2 as (10).

$$\sigma_y^2 = \gamma\sigma_x^2 \quad (\gamma > 0) \tag{10}$$

The error ellipsoid can then be classified by the value of γ. First, we can set $\sigma_x^2 = \sigma_y^2 = \sigma_0^2$ when $\gamma = 1$. Therefore, $\alpha = 45^\circ$ as shown in the following calculation.

$$\begin{aligned} \sigma_u^2 &= \frac{\sigma_0^2 + \sigma_0^2 + \sqrt{4\sigma_{xy}^2}}{2} = \sigma_0^2 + \sigma_{xy} \\ \sigma_v^2 &= \frac{\sigma_0^2 + \sigma_0^2 - \sqrt{4\sigma_{xy}^2}}{2} = \sigma_0^2 - \sigma_{xy} \\ \tan\alpha &= \frac{\sigma_{xy}}{\sigma_0^2 + \sigma_{xy} - \sigma_0^2} \\ &= 1 \end{aligned}$$

If $\sigma_{xy} \sim 0$, the distribution has a circle shape; if it has a large value, the distribution has an ellipsoid shape.

Next, we consider the case of $\gamma \neq 1$ in the equation (10), where we apply the observed data properties to the variables in the equations (7) – (9). Since $\sigma_x^2,\ \sigma_y^2$, and σ_{xy} are at most on the order of 10^{-2} given the experimental environment, the σ_{xy} term is on the order of 10^{-4}. Therefore, we can ignore the σ_{xy} term. The equations (7) – (9) can be approximated by the following equations.

$$\sigma_u^2 = \frac{\sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2}}{2} \sim \sigma_x^2 \tag{11}$$

$$\sigma_v^2 = \frac{\sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2}}{2} \sim \sigma_y^2 = \gamma\sigma_x^2 \tag{12}$$

$$\tan\alpha = \frac{\sigma_{xy}}{\sigma_u^2 - \sigma_y^2} = \frac{\sigma_{xy}}{(1-\gamma)\sigma_x^2} \tag{13}$$

In the situation, considering the range of γ and signum of σ_{xy}, we can predict the following categories. Hence, $L,\ G$ represent local or global reaction, and $+,\ -$ after the L or G represent strong or weak effect. $_,\ ^-$ represent the spreading to lower or upper side of data.

- case $\sigma_{xy} \sim 0$:
  - $\gamma \sim 1$: $L0G0$
    The error ellipsoid distribution has circle shape.
  - $0 < \gamma \ll 1$:*L+G–*
    The shape becomes parallel to the x axis, and $\alpha \sim 0^\circ$.

  - $\gamma \gg 1$: *L–G+*
    The shape becomes parallel to the y axis, and $\alpha \sim 90°$.
- case $\sigma_{xy} > 0$: L_-G^-
  The shape becomes parallel to the x (in case of $0 < \gamma < 1$) or y (in case of $\gamma > 1$) axis, and $0° \ll \alpha < 90°$.
- case $\sigma_{xy} < 0$: L^-G_-
  The shape becomes parallel to the x (in case of $0 < \gamma < 1$) or y (in case of $\gamma > 1$) axis, and $90° \ll \alpha < 180°$. In case of $\sigma_x^2 \sim \sigma_y^2$ $(\gamma \sim 1)$, $\alpha \sim 135°$.

III. Typology of Pupillary Response based on LGR-Map

To evaluate the validity of the LGR-Map designed in Section II, we analyzed the pupillary response. The LGR-Map analysis was conducted using the pupillary response data measured for the controlled narration in the form of Figure 2.

A. Characteristics of Data for Generating LGR-Map

The data used are those obtained by [6]. The data profile is as follows. The narration source used in the experiment is designed as shown in Figure 2.

1) The narration is played back in Japanese, and is a short sentence consisting of about 30 syllables.
2) One ANEW corresponding to either high-positive valence V_{++}, high-negative valence V_{--}, or neutral valence V_N was placed at t_{vs} in one sentence.
3) After the appearance of an ANEW, we assigned an expression that characterizes the mood of the whole sentence as positive (At_+) / neutral (At_N) / negative (At_-).
4) After t_4, the analysis interval from t_a as t_5 to Δt, where the pupillary response is expected to start, was set as analysis interval 1.
5) The response that occurs at $0.5\Delta t$ before and after the end of narration was defined as analysis interval 2.

Therefore, analysis interval 1 was defined as local reaction (instantaneous reaction) and analysis interval 2 as global reaction (integrated reaction). Twenty-one participants in their 20s were included, but data of two participants were excluded due to inaccuracy.

Table I shows the results of subjective evaluation of narration stimuli by participants. The narration stimuli are composed of V_{--}, V_{NN}, V_{++} and At_-, At_N, At_+. The participants listened to each stimulus and then evaluated their impressions on a 7-point scale from high negative to high positive, indicating whether their ratings were consistent with the valence or the atmosphere. However, cases in which the impression matched less than 10 participants were excluded.

B. LGR-Map to Represent Individual Participants' Response Sensitivity

For each participant, an LGR-Map was created for all narration stimuli for the pupillary responses obtained under the above conditions. In order to confirm that the distribution was independent of the size of the individual pupillary response, median-normalized values within analysis interval 1 and analysis interval 2 were used for the plots.

TABLE I. The results of the emotional arousal effect of sentences and the impression evaluation. V denotes valence, At denotes atmosphere, S_I denotes Score of Impression.

V	At	$V = S_I(\%)$	$At = S_I(\%)$	Number of Trial
V_{NN}	At_N	38	38	50
V_{--}	At_-	58	58	54
V_{++}	At_+	54	54	49
V_{NN}	At_-	16	39	24
V_{NN}	At_+	17	54	34
V_{--}	At_+	8	28	29
V_{++}	At_-	4	52	49

TABLE II. The α and flattening rate of the error ellipse in the LGR-Map for the characteristics of the narration stimulus. F_{ee} denotes the flatness.

V	At	α	F_{ee}	V	At	α	F_{ee}
V_{NN}	At_-	63.6°	0.506	V_{--}	At_-	28.4°	0.091
V_{NN}	At_N	43.1°	0.350	V_{--}	At_+	12.3°	0.434
V_{NN}	At_+	32.4°	0.246	V_{++}	At_-	52.2°	0.182
				V_{++}	At_+	−7.43°	0.194

From the equation (1), we expect that the distribution of individual participants' pupillary responses in the LGR-Map can be classified into five types. Figure 3 shows a representative example of an LGR-Map created using the pupillary responses of individual participants to narration stimuli. As shown in Section II, (a) in Figure 3 is $L+G-$, same as (b) is $L0G0$, (c) is L^-G_-, (d) is L_-G^-, and (e) is $L-G+$. When creating the LGR-Map for individual participants, we also examined whether there was a bias in the pupillary response to a particular valence or atmosphere, but no bias was found.

IV. Discussion: Implications of LGR-Map

A. LGR-Map for Characterizing Individual Participant

The classification of individual participants was not characterized by a distinctive response to the combination of (valence, atmosphere), which indicates emotion, suggesting that it was simply determined by the distribution of $w_{ik}(t_j)$, which is indicated by the equation (1). The intensity of $w_{ij}(t_j)$ is considered to change depending on the intensity of the individual's experience of emotion. If the overall experience of emotion is weak, or if the experience of emotion is weak for some reason and almost no emotion is generated, the response of L_-G^- is expected to be shown. When the reaction is triggered by either valence or atmosphere, it is considered to have a reaction of *L+G–* or *L–G+*. If the reaction is equally distributed between valence and atmospheres, the reaction is considered to be L^-G_-. If the reaction is completely random, it is considered to be $L0G0$.

B. LGR-Map for Categorizing Narrations

Next, we consider human responses to ANEWs used as narration stimuli. Since ANEWs are basically emotion references elicited when people hear the word, we believe that it is possible to evaluate the validity of narration stimuli that show the same atmospheres as ANEWs by using the LGR-Map type classification.

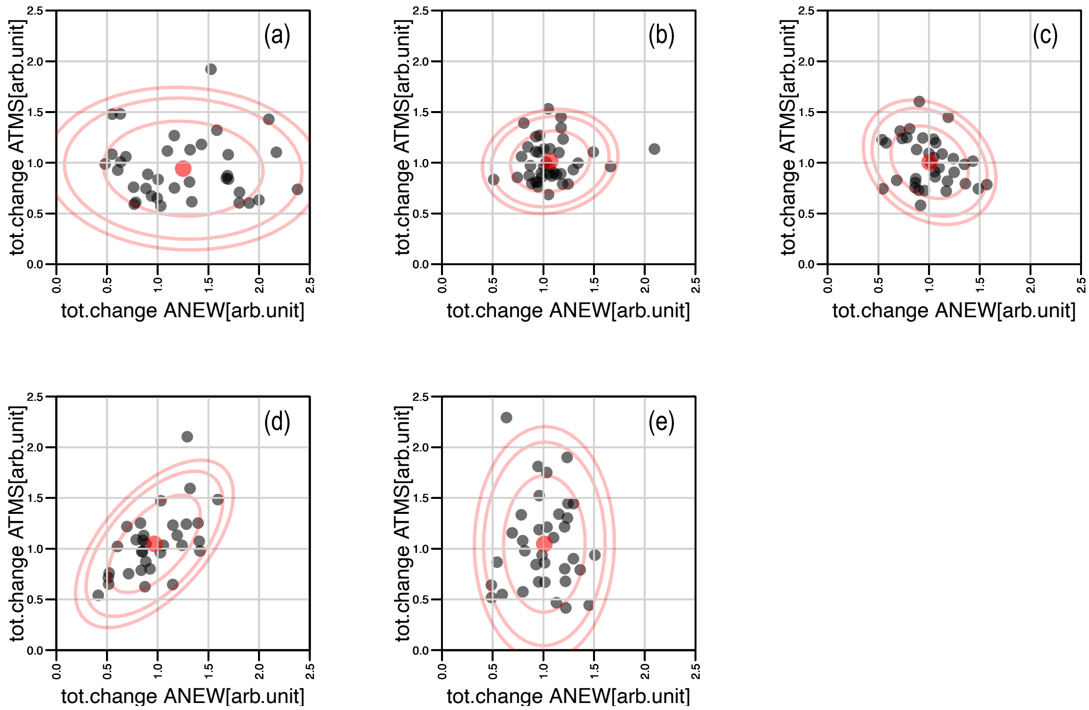

Figure 3. Examples of LGR-Maps for individual participants. The LGR-Map an individual participant is normalized by a median of r_{all} near ANEW and near the end of the sentence, respectively. The oval lines indicate 66%, 90%, and 95% confidence levels from the inside. The categories in LGR-Map are as follows: (a) $L + G-$, (b) $L0G0$, (c) $L^{-}G_{-}$, (d) $L_{-}G^{-}$, (e) $L - G+$.

Table II shows the values of α and flattening F_{ee}, which are the features of LGR-Map. The features in the LGR-Map are created by combining the valence and atmospheres into 9 patterns. There were three responses to each stimulus pair. Trials with fewer than 10 trials showing the level of response to a stimulus pair were excluded from the analysis, considering them to be less significant even if an error ellipse was written.

The table shows the following characteristics. For $V_{--}At_{-}$, F_{ee} is almost zero, indicating that it is a circular distribution. Therefore, (a) is classified as $L0G0$. For $V_{--}At_{+}$, α is 13.2°, almost parallel to the x axis, so it is classified as $L + G-$. For $V_{NN}At_N$, it is classified as $L_{-}G^{-}$, because $\alpha \sim 45°$. $V_{NN}At_{-}$, $V_{NN}At_{+}$ are not certain because the value of α is ambiguous, but we can classify them as $L_{-}G^{-}$ for the reason described later. Since $V_{++}At_{-}$, $V_{++}At_{+}$ have ambiguous α values and F_{ee} values are not circular, we cannot indicate which type they can be classified into at this time.

From the above, the following possibilities are considered for $V_{--}At_{-}$, $V_{--}At_{+}$, $V_{NN}At_N$. For $V_{--}At_{-}$, values of valence and atmosphere have negative each. In this case, valence and atmosphere are the same characteristics, so that we anticipate that participants' pupillary responses are almost uniform as indicated by Murakami et al. [4][5]. Taken together, these results suggest that the distribution of the LGR-Map is random, centered on a representative LGR-Map value.

For $V_{--}At_{+}$, the response is negative valence, with positive atmosphere. This, together with $V_{--}At_{-}$, can be interpreted as follows. The response to the negative valence was scattered, but the response to the pupillary response in the atmospheres was reasonably consistent, resulting in the values in the y axis being almost consistent and the distribution in the x axis being broadened. This is thought to be due to the broadening of the x-axis distribution.

Next, we consider the case of $V_{NN}At_N$. Both valence and atmosphere were neutral, that is no emotion is induced. It indicates that no matter where in the instantaneous or integrated area the pupillary response is measured, no change in emotion occurs for the same person or narration. Therefore, both pupillary responses show almost similar values, which is a good sign that the distribution is close to a straight line with $y = x$.

V. Generated Emotion from Sentences with Multiple ANEWs

The LGR-Map introduced in Section II is based on the relationship between a single ANEW and sentence-final expressions (positive/negative). In this case, since the factors that promote emotional change are either emotion-inducing words or sentence ends, the LGR-Map can be created by examining the degree of total change in respective pupillary responses.

However, in the real world, it is not uncommon to find multiple ANEWs in a single sentence, or to induce emotions from information given by a sequence of multiple sentences. In order to adapt LGR-Map to such a situation, several modifications may be necessary for the definition of the LGR-Map. In this section, as a first step, we discuss the total change of pupillary response expected when auditory stimuli consisting of one or two sentences with two ANEWs are presented to a participant. In addition, the degree of memorization of the given stimuli is also discussed with reference to Figure 1.

A. Basic Design

The emotion induction when participants hear sentences with multiple ANEWs is affected by various among ANEWs, compared to the emotion induction with a single ANEW. The interference is caused by the valences of ANEWs or the presentation interval between them. In this subsection, we discuss the effects of each type of interference.

Table III shows the classification of emotion induction by the valences of two ANEWs, $ANEW_1$ and $ANEW_2$, vis-a-vis the interval between them, and predicted degree of memorization for the content of stilumi. The valences V and V' are defined as follows:

$$\begin{aligned} V,\ V' &\in \{V_{++},\ V_{--}\}, \\ \text{where } (V,\ V') &= (V_{++},\ V_{--}) \text{ or } (V_{--},\ V_{++}). \end{aligned}$$

We set the interval between $ANEW_1$ and $ANEW_2$, i.e. T_{int} as short or long.

Depending on the combination of possible values of valence for ANEWs, the following four cases are possible for emotion induction and the resulting degree of memorization of the concept carried by the presented stimuli.

(a) ANEWs valence: $(V,\ V)$, T_{int}: short
After the short T_{int} from the presentation of $ANEW_1$, $ANEW_2$ appears, which means emotionally similar concepts to $ANEW_1$ are triggered in close proximity to the presentation of $ANEW_2$. The concepts carried by the stimuli's contents might be highly activated as a result of triggering the adjacent regions in long-term memory, which have overlapping areas of the presented two ANEWs, at short time intervals. This would result in the strengthening of the association between the concepts and the ANEWs. As a result, the concept, i.e., the stimuli's contents, will be recalled easily, and can be answered easily on the recall test.

(b) ANEWs valence: $(V',\ V')$, T_{int}: long
After the long T_{int} from the presentation of $ANEW_1$, $ANEW_2$ appears, which means emotionally similar concepts are triggered at a long interval. The concepts carried by the presented stimuli will be activated in isolation due to the quasi-independent triggering of long-term memory. Even if the triggered areas have overlapping areas, it is not likely that they reinforce each other because the timing of activation is not synchronized. The performance for the recall test will be limited because the mutual reinforcement of activities of the concepts of input stimuli would not be as effective as in the case of (a), in which the presentation of succeeding ANEWs in a short-interval assures the possibility of a synchronized activation of the concepts.

* The following two cases are incongruent versions of (a) and (b). The two ANEWs trigger isolated regions of long-term memory when presented. It is not likely that they jointly reinforce the activation of long-term memory. However, the short or long interval between the two ANEWs would have distinctive effects on memory test as described below.

(c) ANEWs valence: $(V,\ V')$, T_{int}: short
After the short T_{int} from the presentation of $ANEW_1$, $ANEW_2$ appears, which means emotionally different concepts from $ANEW_1$ are triggered in close proximity to the presentation of $ANEW_2$. The concepts carried by the presented stimuli will be activated as a result of triggering long-term memory, which might not have overlapping areas, at short time intervals. This would result in a weak association between the concepts and the ANEWs. As a result, the concept, i.e., the stimuli's contents, will be hard to recall, and can not be answered easily on a recall test.

(d) ANEWs valence: $(V',\ V)$, T_{int}: long
After the long T_{int} from the presentation of $ANEW_1$, $ANEW_2$ appears, which means emotionally different concepts from $ANEW_1$ are triggered at a long interval. The concepts carried by the presented stimuli will be activated independently as a result of triggering of long-term memory with possibly the least overlapping areas of the two ANEWS. As a result, the performance of the recall test does not improve, and the memory recall depends on the degree of past experience.

Among these states, we predict that interference for association strength between ANEWs is expected in the behavior of r_{all}, especially for cases (a) and (c), i.e., when ANEWs have a short appearance interval. We describe r_{all} for $ANEW_1$, $ANEW_2$, and *ES* by $r_X(X = \{ANEW1/ANEW2/ES\})$; we regard r_X as a quantity that indicates the degree of response to $ANEW_1$/$ANEW_2$/*ES*. The relationship between r_X's in terms of the degree of their amount indicates which partial ANEWs or whole sentences should have caused a stronger response. Based on this assumption, we extend the LGR-Map introduced in Section II. Toward constructing the extended LGR-Map, we attempt to find the influence of emotion induction as follows: to see the local response, we calculate the difference of r_{ANEW1} and r_{ANEW2}; for global response, we calculate the difference of r_{ANEW1} or r_{ANEW2} and r_{ES}.

Figure 4 shows the relationship between the profile of r_{PD}, i.e., numerical characteristics of r_X, and the characteristics on the extended LGR-Map. Figure 4 (A) represents the possible observed pupillary response curve, which is derived from emotion induction response. The red/blue/green rectangle areas in Figure 4 (A) represent the area of the total change in pupillary response to the first ANEW ($ANEW_1$) / second ANEW ($ANEW_2$) / the end of sentence (*ES*), $r_{ANEW1/ANEW2/ES}$, respectively. $R1/R2/RS$ in Figure 4 represents the responses to $ANEW_1$/$ANEW_2$/*ES*. In this case, there are six different combinations of the two quantities between r_Xs, as shown in Table IV. Figures 4 (B) and (C) represent the behavior

TABLE III. CLASSIFICATION OF EMOTION INDUCTION BY THE VALENCES, $ANEW_1$ AND $ANEW_2$, AND THE INTERVAL BETWEEN ANEWS AND PREDICTION OF THE DEGREE OF MEMORIZATION

Valence of $ANEW_1$	Interval T_{int}	Valence of $ANEW_2$: V — degree of related activation of concepts	Valence of $ANEW_2$: V — degree of memorization	Valence of $ANEW_2$: V' — degree of related activation of concepts	Valence of $ANEW_2$: V' — degree of memorization
V	short	high	high	dependent on concepts	dependent on concepts
V'	long	low (hard to associate)	memorization to the extent of past experience	low (hard to associate)	low

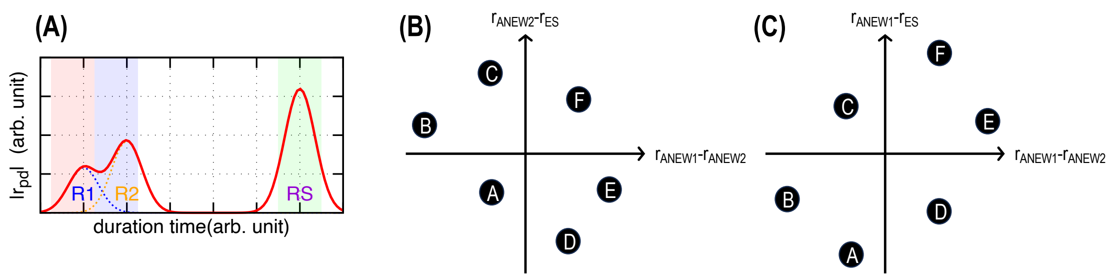

Figure 4. Observed images of the magnitude of the absolute total pupil response for $ANEW_1$, $ANEW_2$, and ES. (A) shows the composite of $|r_{pd}|$ for $ANEW_1$, $ANEW_2$, ES. (B) and (C) projected onto the extended version of the LGR-Map plane, which shows the behavior between r_X corresponding to Table IV.

for different r_X's in the extended LGR-Map, which are categorized in Table IV. Figures 4 (A) through (F) indicate the types based on the relationship between r_Xs, as shown in Table IV.

We considered that the degree of valence induced in conjunction with concepts, as predicted in Table IV, leads to emotion induction, which is directly related to the magnitude of r_X. Compared to the case where a single ANEW appears in isolation, the consideration shown in Table III for two ANEWs suggests that the following is expected: The presentation of two congruent ANEWs will affect each other in a direction that is more likely to induce emotion. A short interval presentation will, in particular, induce a particularly strong emotion induction. The concepts are then recalled under the influence of stronger association strength compared to weak emotion induction. As a result, the results of the recall test will improve. This situation can be rephrased as follows. Once an emotion is induced by $ANEW_1$, the emotion induction is continued by the same type of $ANEW_2$ shortly thereafter. Since $ANEW_2$ continues the emotion induction while the emotion was once induced by $ANEW_1$, at least the effect of local ANEWs is expected to be $r_{ANEW1} < r_{ANEW2}$. From this, we can specify that one of the types A∼C in Table IV is advantageous. The relationship between $ANEW_1$/$ANEW_2$ and ES depends on the relationship with r_{ANEW1}/r_{AEW2}, but it is reasonable to assume that r_{ES} will have a smaller value than r_{ANEW1}/r_{AEW2} under the current assumptions after stronger emotions are generated. Based on the above, the behavior of r_X in Table III (a) is expected to be consistent with the behavior shown in Table IV (B) and (C).

B. Preliminary Experiment

To confirm the prediction considered in Section V-A, we performed a preliminary experiment. The narration structure

TABLE IV. BEHAVIOR BETWEEN r_X WHEN LGR-MAP IS EXTENDED.

	X-axis: $r_{ANEW1} - r_{ANEW2}$	Y-axis: $r_{ANEW2} - r_{ES}$ Figure 4(B)	Y-axis: $r_{ANEW1} - r_{ES}$ Figure 4(C)
A	$-$	$-$	$--$
B	$--$	$+$	$+$
C	$-$	$++$	$+$
D	$+$	$--$	$-$
E	$++$	$-$	$+$
F	$+$	$+$	$++$

of the auditory stimuli designed in Figure 2 was modified as shown in Figure 5, and 32 auditory stimuli containing two ANEWs were prepared according to the following policy.

- We adopt ANEWs which arousal values are moderate and valence values are V_{++} with the range of $(7.70 - 9.00)$ or V_{--} with the range of $(1.00 - 1.99)$.
- Each sentence, i.e., an auditory stimulus, consists around 65 syllables.
- The interval between two ANEWs is as follows:
  - Each ANEWs are almost adjacent to each other, and has an inference on the pupillary response to ANEW (we adopt the 1 second interval for this experiment).
  - The two ANEWs are apart long enough to have little interference with each other on the pupillary response to ANEWs (we adopt the 7 second interval for this experiment).

In addition, a recall test was conducted on the content of every eight sentences to check the degree of memorization.

C. Result of Preliminary Experiment

The analysis of the preliminary experiment was conducted as follows. Because this was a preliminary experiment, there were two participants. Since there are two ANEWs in each auditory stimulus, we need to extend the LGR-Map introduced

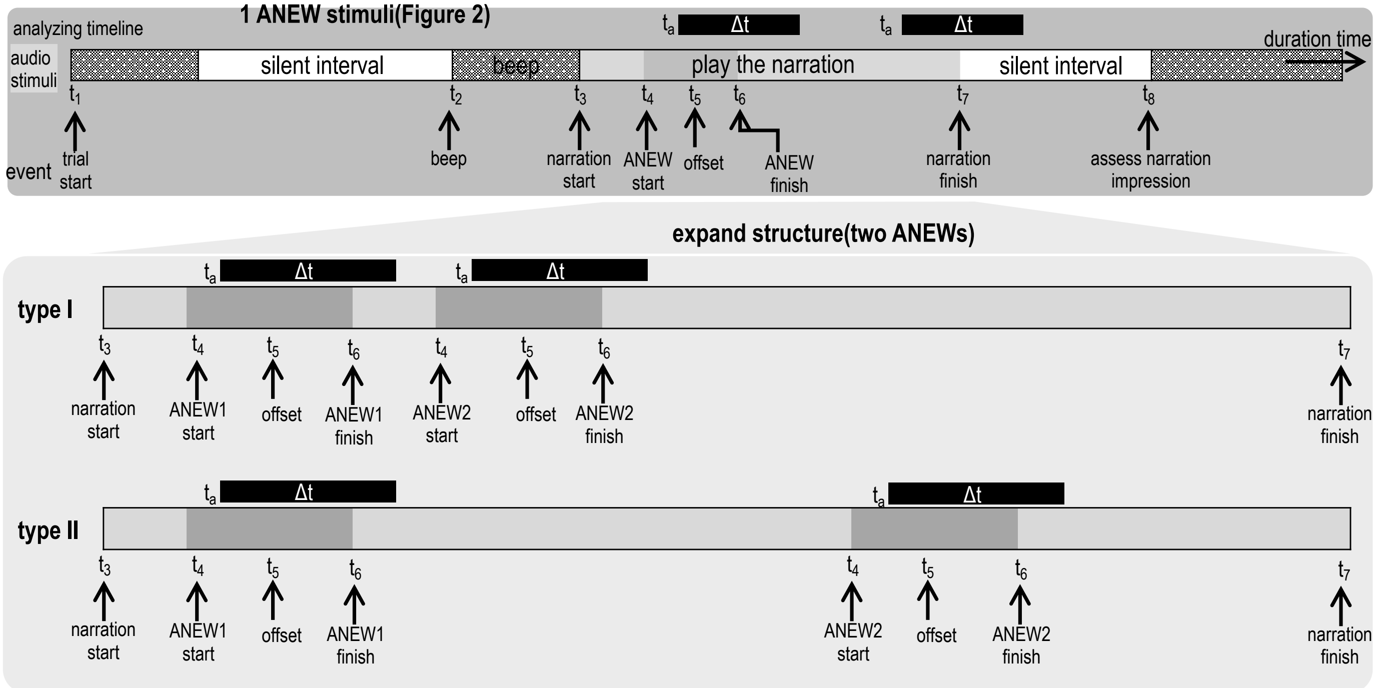

Figure 5. Timeline of presented stimuli including multiple ANEWs.

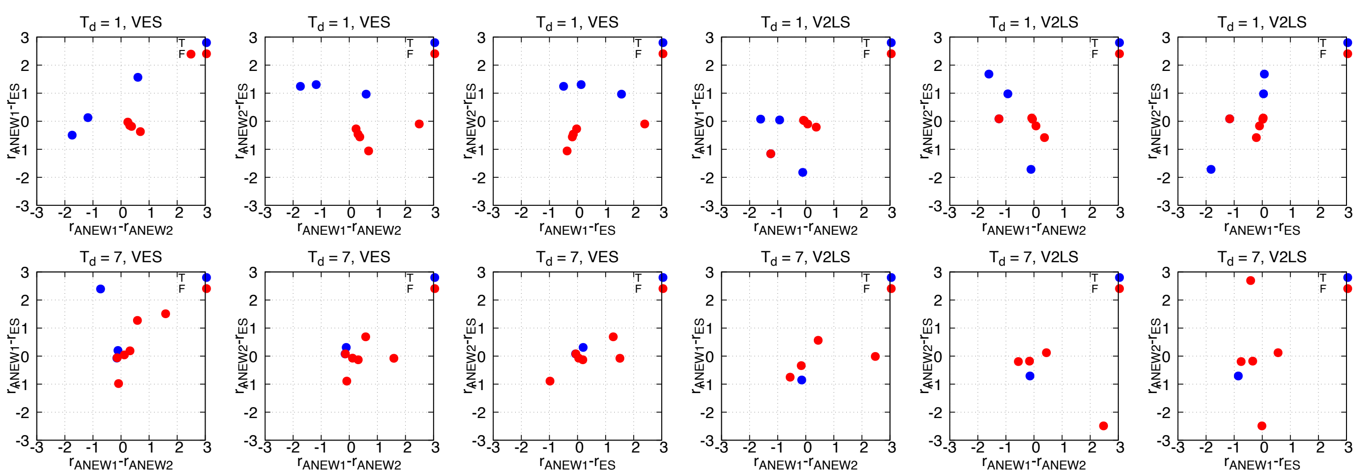

Figure 6. Measurement results between r_X. The horizontal and vertical axes are the differences between the observations in the combinations shown in the figure. Blue marks indicate that the corresponding presented stimulus was memorized, and red marks indicate that it was not memorized.

in Section II to capture the trend characterized by the local and global reactions. We devised the following two methods as candidates for a revised LGR-Map to examine its suitability for capturing trends. The purpose of the analysis is to estimate the appropriateness of the experimental design for defining the extended LGR-Map.

- Method 1: The graph was modified by taking $r_{ANEW1} - r_{ANEW2}$ on the horizontal axis to see the degree of the local impact for a pupillary response, and $r_{ANEW1} - r_{ES}$ and $r_{ANEW2} - r_{ES}$ on the vertical axis to see the impact of ANEW on the global pupillary response.
- Method 2: To directly assess the local-global influence, use $r_{ANEW1} - r_{ES}$ as the horizontal axis and $r_{ANEW2} - r_{ES}$ as the vertical axis.

In order to see the influence of ANEWs and induced emotion in each auditory stimulus, the data were classified based on the difference between the judged valence score S_p by participants for each stimulus and the valence $(V_{1/2})$ for each ANEWs in the stimuli. Some of the results are shown in Figure 6.

In describing the graphs, the relationship between the valence of ANEWs and S_p in auditory stimuli was checked as follows, and graphs were described for each category as listed below.

TABLE V. SUMMARY OF FEATURE FOR FIG.4, AND THE RESULTS OF THE RECALL TEST FOR AUDITORY STIMULI. THE RESULTS OF RECALL TEST ARE SUMMARIZED BY THE NUMBER AND PERCENTAGE OF THE AUDITORY STIMULI, WHICH ARE CATEGORIZED BASED ON FIGURE 4 TYPES, WHICH ARE MEMORIZED (MEM. TRUE) OR ARE NOT MEMORIZED (MEM. FALSE).

Type	$r_{ANEW1} - r_{ANEW2}$	$r_{ANEW2} - r_{ES}$	$r_{ANEW1} - r_{ES}$	Feature	Mem. True	Mem. False
A	$-$	$-$	$--$	$r_{ANEW1} < r_{ANEW2} < r_{ES}$	0.10(6)	0.08(5)
B	$--$	$+$	$+$	$r_{ANEW1} < r_{ES} < r_{ANEW2}$	0.05(3)	0.07(4)
C	$-$	$++$	$+$	$r_{ES} < r_{ANEW1} < r_{ANEW2}$ tend to mem.	***0.10***(6)	0.07(4)
D	$+$	$--$	$-$	$r_{ANEW2} < r_{ANEW1} < r_{ES}$ hard to mem.	0.07(4)	***0.19***(11)
E	$++$	$-$	$+$	$r_{ANEW2} < r_{ES} < r_{ANEW1}$ hard to mem.	0.02(1)	*0.08*(5)
F	$+$	$+$	$++$	$r_{ES} < r_{ANEW2} < r_{ANEW1}$ hard to mem.	0.05(3)	***0.12***(7)

$$\begin{aligned}
\text{VES}: &\quad |V_1 - S_p| < 1,\ |V_2 - S_p| < 1 \\
\text{V1ES}: &\quad |V_1 - S_p| < 1,\ |V_2 - S_p| \geq 2 \\
\text{V1LS}: &\quad 1 \leq |V_1 - S_p| < 2,\ |V_2 - S_p| \geq 2 \\
\text{V2ES}: &\quad |V_1 - S_p| \geq 2,\ |V_2 - S_p| < 1 \\
\text{V2LS}: &\quad 1 \leq |V_2 - S_p| < 2,\ |V_1 - S_p| \geq 2 \\
\text{N/A}: &\quad |V_1 - S_p| \geq 2,\ |V_2 - S_p| \geq 2
\end{aligned}$$

Although there are 64 data points in total, only the categories that contain enough data for evaluation are shown, excluding those with a small number of data per category. Since the recall test was conducted at the same time, the r_X data for the memorized stimuli are shown in blue dots, and the r_X data for the not memorized stimuli are shown in red dots. The results suggest that the interval of 1 second between $ANEW_1$ and $ANEW_2$ and the VES category are the easiest to separate the memorized from not memorized stimuli.

Analysis of these results in comparison with Figure 4 leads to the following conclusions:

- $r_{ANEW1} - r_{ANEW2} < 0$, which means $r_{ANEW1} < r_{ANEW2}$,
- $r_{ANEW1} - r_{ES} \leq 0$, which means $r_{ANEW1} < r_{ES}$,
- $r_{ANEW2} - r_{ES} > 0$, which means $r_{ES} < r_{ANEW2}$.

Integrating all of them, we find that the situation might correspond to the situation defined in Table IV B or C. That is, the emotion initially generated by $ANEW_1$ is reinforced by $ANEW_2$ having the same type of valence, and the sentence presented in audio comes to an end before the emotional effect is fully extinguished.

Table V summarizes the results for all data. Column 1 in the table shows the graph type of Table IV, columns 2 to 4 show the positive and negative difference of r_X, and column 5 shows the relationship between r_X. The number of provided stimuli that were memorized was counted for each category of Table IV, and those that were memorized were described in column 6, Mem. True, and those that were not memorized were described in column 7, Mem. False. As a result, the state of Table IV C has a high degree of memorization, and the state of D or F has a low degree of memorization. The result indicates as follows:

$ANEW_1$ and $ANEW_2$, i.e., the distance between ANEWs in a provided auditory stimulus and the end of a sentence, should not be close together but should have a certain distance between them for easier memorization. ANEWs should include those with the same kind of valence as much as possible. One possibility is that the proximity of ANEWs generates a controlled valence, which is maintained for a long time if it exceeds a certain level of intensity, and that this valence, together with the concepts entered with the ANEWs, is stored and reinforced in the long-term memory. This is regarded to correspond to short-term learning. Therefore, in the recall test, the most recently learned items are presented in the order in which they were most recently learned, leading to improved memorization output.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we focused on the pupillary response and proposed the LGR-Map as a classification method for pupillary changes associated with emotional changes. The LGR-Map indicates whether an individual's pupillary response to a stimulus is more likely to respond to local informaion or contextual information.

In order to propose the LGR-Map, we needed a cognitive model that describes how people's emotions are induced in response to stimuli from the outside world. Thus, we constructed a model of human emotional responses based on the CI-model. We assumed that the input stimuli have the feature of auditory information, that is, transient information. At each time point, the auditory information was captured as a single packet of sound waves, and each packet was matched with the long-term memory only once. ANEW was assigned to one of the packets and assumed to appear somewhere in the sentence. We assumed a situation where the valence of the whole sentence is determined by the sentence-final expression. For each of them, we considered an emotional response appeared based on the CI-model. When considering the above situations, we thought that there was some kind of pupillary response for each emotional reaction. We described the pupillary response to ANEW as "local" response, and the pupillary response to the end of a sentence as "global" response. In this case, the local and global pupillary responses can be represented in a two-dimensional plane. Based on this idea, we proposed the LGR-Map. The shape classification of the LGR-Map was based on the variance of the error ellipsoid. The results indicated that the LGR-Map could be classified into five types according to the covariance of the local and global pupillary responses and the trend of the dispersion of the local pupillary response.

Based on a series of ideas, 36 auditory stimuli with various characteristics embedded in ANEWs and sentence-final expressions were actually given to 19 participants, and LGR-Maps were created. As a result, we confirmed that five types of shapes were recognized for individual pupillary responses. The effectiveness of the LGR-Map was confirmed by the fact that five different shapes were found for individual pupillary responses in 19 participants who were given 36 auditory stimuli with various characteristics embedded in ANEWs and sentence-final expressions.

As an extension of the LGR-Map, an application to the pupillary response in the case of multiple ANEWs was also considered and preliminary experiments were conducted. The results of preliminary experiment indicated the consistency with the Figure 1. Further analysis and experiment of auditory stimuli with multiple ANEWs will indicate further insights.

Image language plays a greater role than symbolic language in real-time communication. However, memes, which are words, play a major role in transmitting and accumulating knowledge over a long period [16]. Knowledge represented by a network, a meme, or a symbolic node, develops links to image nodes associated with it. Image nodes are formed in response to an individual's actual perceptual, cognitive, and motor experiences, represent something unique to each individual. This is very different from symbolic nodes, which are shared within a single culture. Communication through memes (words) is a form of communication through language nodes that aggregate a large amount of information, allowing for the exchange of a large amount of information with a small amount of information (words). This is achieved through the activation of image nodes that are associated with the language node. However, it is not guaranteed that the associated image nodes centered on the language node are consistent on both sides of the interlocutor. Therefore, errors in the transmission of information due to this discrepancy are inevitable [17].

For example, the proliferation of the Social Networking System (SNS) allows transmission errors to be amplified in an extremely short period of time. While considering these characteristics of SNS, it is necessary to establish a method to realize verbal communication that does not cause transmission errors and to build a community where people can communicate healthily. Toward this end, an approach that focuses on the activation of knowledge centered on language nodes, as shown in this study, is promising.

The application of the LGR-Map will make it possible to provide adaptive content for individuals. The method of implementation will be an issue in the future.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 19K12246 / 19K12232 / 20H04290 / 22K12284 / 23K11334 , and National University Management Reform Promotion Project. The authors would like to thank Editage (www.editage.com) for English language editing. MH also wants to thank to Nagai N · Promotion Foundation For Science of Perception for their finantial support.

References

[1] K. Nakahira, M. Harada, and M. Kitajima, "Local-global reaction map: Classification of listeners by pupil response characteristics when listening to sentences including emotion induction words -toward adaptive design of auditory information," in COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, 06 2023, pp. 79–85.

[2] N. Vallez et al., "Automatic museum audio guide," Sensors, vol. 20, no. 3, 2020. [Online]. Available: https://doi.org/10.3390/s20030779

[3] R. Hirabayashi, M. Shino, K. T. Nakahira, and M. Kitajima, "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 (VISIGRAPP 2020), vol. 2, 2020, pp. 162–169.

[4] M. Murakami, M. Shino, K. T. Nakahira, and M. Kitajima, "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 (VISIGRAPP 2021), vol. 2, 2021, pp. 89–100.

[5] M. Murakami, M. Shino, M. Harada, K. T. Nakahira, and M. Kitajima, "Effects of emotion-induction words on memory and pupillary reactions while viewing visual stimuli with audio guide," in Computer Vision, Imaging and Computer Graphics Theory and Applications, A. A. de Sousa et al., Eds. Cham: Springer International Publishing, 2023, pp. 69–89.

[6] S. Moriya, K. T. Nakahira, M. Harada, M. Shino, and M. Kitajima, "Can pupillary responses while listening to short sentences containing emotion induction words explain the effects on sentence memory?" in Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2023, Volume V: HUCAPP, Online Streaming, February 19-21, 2023. SCITEPRESS, 2023, pp. 213–220.

[7] J. Z. Lim, J. Mountstephens, and J. Teo, "Emotion recognition using eye-tracking: Taxonomy, review and current challenges," Sensors, vol. 20, no. 8, 2020. [Online]. Available: https://doi.org/10.3390/s20082384

[8] L. Shu et al., "A review of emotion recognition using physiological signals," Sensors, vol. 18, no. 7, 2018. [Online]. Available: https://doi.org/10.3390/s18072074

[9] W. Kintsch, "The use of knowledge in discourse processing: A construction-integration model," Psychological Review, vol. 95, 1988, pp. 163–182.

[10] W. Kintsch, Comprehension: A paradigm for cognition. Cambridge, UK: Cambridge University Press, 1998.

[11] C. Wharton and W. Kintsch, "An overview of construction-integration model: A theory of comprehension as a foundation for a new cognitive architecture," SIGART Bull., vol. 2, no. 4, jul 1991, p. 169–173. [Online]. Available: https://doi.org/10.1145/122344.122379

[12] J. Russell, "A circumplex model of affect," Journal of Personality and Social Psychology, vol. 39, 12 1980, pp. 1161–1178.

[13] K. T. Nakahira, M. Harada, and M. Kitajima, "Analysis of the relationship between subjective difficulty of a task and the efforts put into it using biometric information," in Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 2: HUCAPP, Online Streaming, February 6-8, 2022, A. Paljic, M. Ziat, and K. Bouatouch, Eds. SCITEPRESS, 2022, pp. 241–248. [Online]. Available: https://doi.org/10.5220/0010906800003124

[14] P. Jerčić, C. Sennersten, and C. Lindley, "Modeling cognitive load and physiological arousal through pupil diameter and heart rate," vol. 79, no. 5, pp. 3145–3159. [Online]. Available: https://doi.org/10.1007/s11042-018-6518-z

[15] M. M. Bradley, L. S. Miccoli, M. A. Escrig, and P. J. Lang, "The pupil as a measure of emotional arousal and autonomic activation." Psychophysiology, vol. 45, no. 4, 2008, pp. 602–607. [Online]. Available: https://doi.org/10.1111/j.1469-8986.2008.00654.x

[16] D. C. Dennett, From Bacteria to Bach and Back: The Evolution of Minds. W W Norton & Co Inc, 2 2018.

[17] M. Kitajima, M. Toyota, and J. Dinet, "How Resonance Works for Development and Propagation of Memes," International Journal on Advances in Systems and Measurements, vol. 14, 2021, pp. 148–161.