Active Long Fixation Correlates with the Formation of Long-Term Memory

적극적 장기 응시와 장기 기억 형성과의 관련성 연구

2015 년 2 월

서울대학교 대학원

협동과정 인지과학전공

김 진 화

공학석사학위논문

Active Long Fixation Correlates with the Formation of Long-Term Memory

적극적 장기 응시와 장기 기억 형성과의 관련성 연구

2015 년 2 월

서울대학교 대학원

협동과정 인지과학전공

김 진 화

Active Long Fixation
Correlates with the Formation of
Long-Term Memory

적극적 장기 응시와 장기 기억 형성과의 관련성 연구

지도교수 장 병 탁

이 논문을 공학석사 학위논문으로 제출함

2014 년 11 월

서울대학교 대학원

협동과정 인지과학전공

김 진 화

김진화의 공학석사 학위논문을 인준함

2014 년 12 월

위 원 장	김 청 택	(인)
부위원장	장 병 탁	(인)
위 원	박 주 용	(인)

Abstract

The application of eyewear accelerates the study on the eye movement, for the eye movement is a non-invasive and convenient indicator of the brain activities. We investigate the eye movements of the subjects watching the kids video. We analyze the video sequences by classifying them into two different sequence groups have the long and short fixation duration, respectively. First, we conduct the long-term memory test of whether fixation duration correlates with long-term memory. Second, we classify its visual constraints into Alert and No Alert types. As a result of the test, the fixation duration itself is not decisive, however, the long fixations which are actively engaged with Alert type movie clips statistically have higher scores, while the short fixations do not. Finally, we propose a simple computational model using the linear regression of two significant features, saliency scrutiny and fixation duration. It may provide an explanatory way to the efficient memory mechanism for the life-long sequences.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The brain is the most intelligent organ of a living thing. It receives many different forms of sensory information and processes this information appropriately with regard to its survival and reproduce. Especially, the visual information takes a very special position among other kinds of sensory information as it does not need to sense a source directly but is transfered to a remote target far freely than any other types of the sense. Furthermore it gives more chance to survive in the situation of being threaten by the predators using the eyes to detect the foes in the remote place before their approaching.

But this notable advantage is not freely given one as it requires more delicate and clever way of interpreting the visual information. Because the visual information can easily be affected by the moment-to-moment environmental changes, heuristic but robust compensation strategies are demanded. Hence, how the brain processes the visual information provides the profound way to study the mechanism that the brain precisely and efficiently processes the most dynamic and enormous sensory information.

There are a lot of studies on the computational modeling for the visual information, which include the visual fragment completion, the scene or object classification and recognition [1, 2], and object tracking [3]. These research topics often tend to focus on the objective for each task, not on the implementation of the method how the brain deals with. As a result, the computational approach to the modeling for the visual information processing of the brain is gradually changed to the optimization problem, which hinders the understanding of the human-level information processing abilities.

Particularly, the object tracking seems to describe how we pay attention to an interesting object, however, the eye movement, mostly controlled by the oculomotor system, is complicated with how the brain works for the acquisition of the visual information [4]. For instance, in the fixation state, the human eyes only recognize the small portion of the whole sight. If you read this paper from an 8-inch distance fixing on one particular letter, you cannot read outside of next two words or about ten letters which are presented in the para-fovea. For the brain is well-known for its parallel processing on the neural circuits, this sequential notion of eye movement for the visual system would be inefficient for information processing. Moreover, we should notice that the experimental results of covert attention and shifting receptive field are also important subject to study [5].

The studies on the reading eye movement, which are relatively well studied by psychologists and neuroscientists [6, 7], reveal that fixation duration is related to the presence of the cognitive process [8]. For an instance, observing its correlation with linguistic attributes [9, 10].

There is a different aspect in studying the eye movement for video stimuli. The duration of fixation is more constrained by the affective content, such as emotional response, including a context of that, compared to the reading

materials. As well as the selection of the next fixation and the direction of a saccadic movement tend to be more liberal than the dominance of horizontal searching of reading does.

Therefore, the study of the eye movement on the video stimuli has been neglected due to the research complexities and the methodological difficulties [11]. However, the recent advancement of the sensory device, like the Google Glass and the mobile devices for eye tracking, promotes studies on the video stimuli and in natural experimental environment, to apply the research models or applications on mobile devices.

We investigate the characteristics of eye movements on the video stimuli focusing on the formation of long-term memory using recognition test. The basic elements of the eye movements are segregated into the fixation duration and the saccade vector, which consists of the saccade direction and the length of the saccadic movement [12, 13]. In this study, we focus on the characteristics of the fixation duration as the evidence of the cognitive process to memorize. Moreover, as the movie clips which potentially induce emotional arousal are known to increase the recognition of the previously watched movie clips [14, 15], we will see if the arousal effect is asserted by the duration of fixation. Finally, we propose a linear regression model based on the active features of eye movement, saliency scrutiny and fixation duration.

Chapter 2

Materials and Methods

2.1 Study 1

For this study, we prepared the video material *Pororo Season 3*, which is a famous kids video in Republic of Korea. In this video, there are artificial 3D-rendered characters and other objects who have distinctive traits, so, visual features are easily captured by eye movements. *Pororo Season 3 DVD 1* contains 13 consecutive episodes, each with a different single storyline. The total playing time is 67 minutes and 50 seconds.

We recruited 18 participants with normal vision (11 males, 7 females; 23-31 age), who are graduate students in Seoul National University, voluntarily participated in the study. For the stability of collecting data and the eye gear requires appropriate head circumference, we chose the target participants for these studies. All participants had not experienced a brain damage or a behavioral disorder. The participants were first time viewers of the video, *Pororo Season 3*. To prevent distraction, each participant took a set of tests for the

two-split video, one is about 32 minutes of the first half and the other is about 36 minutes of the later half, each on the other day. Later then, we merged two parts into one manually, in a way that the results do not overlap.

Participants watched the kids video in the room which has the experimental settings. The room is about 3 square meters surrounded by the opaque curtains. On the side of the room, a wide-screen HDTV (1920x1080 resolution, 885 mm x 500 mm, 16:9 ratio) was installed, and 2.1 channel speakers. Participants were guided to sit on comfortable sofa in front of 1.7 m from the TV screen.

Concurrently, the eye movements and the user-perspective scenes were recorded by *Tobii Glasses eye tracker*, the corneal reflection based system with a sampling rate of 30 Hz. We used the *I-VT algorithm* as the fixation filter (system default), which classified fixations with the velocity threshold of 30 degree per second. Usually, the saccadic eye movements are discriminated with low velocities (less than 100 degree/second) and high velocities (higher than 300 degree/second), so the velocity-based classification is simple but a reasonable approach [16].

We classified the event types of the eye movements into three categories; fixation, saccade and unclassified. The unclassified data were discarded and not used in this study.

2.2   Study 2

For 11 participants who have participated in *Study 1*, we prepared the controlled memory test for each participant. We conducted this study 3-4 months after the *Study 1* (the intervals are not consistent due to schedule conflicts). A memory test consisted of total 20 movie clips; 8 movie clips for long fixations, another 8 movie clips for short fixations, and the remaining 4 movie clips for the control, which were not seen in the previous study. In detail, the lengths of all movie

clips were each 3-second long. *The long fixation sequences* were randomly picked from each participant's data containing the fixation longer than 1400 ms in the middle of the movie clip. *The short fixation sequences* were randomly picked from each participant's data containing the fixation shorter than 300 ms. The 4 control movie clips were randomly picked from the other season of *Pororo* series, *Pororo Season 2*.

Each participant identified 20 movie clips that were sorted randomly. Each participant gives the movie clip the score between 1 and 5, which is an integer, depending on the assurance of whether he or she saw the movie clip before or not. Notice that, in this study, 1 is the lowest score, not 0, which means that the movie clip is surely not seen.

Chapter 3

Fixation Duration

Fixation duration takes a important position in reading task, for the durations of eye fixations seems to be constrained by the linguistic features of the fixated word [10, 9]. In the video watching task, we anticipated the characteristics of fixation duration are different from those in the reading task. As we expected, the fixation duration was changed more drastically, up to 10 seconds during watching the video. These changes were partly caused by the sequential changes of the visual stimuli and the fluctuated responses of the oculomotor system and cognitive processes. The length of fixation duration is not a deterministic property, because more than a single component dynamically contribute to the final motor command and the execution for the oculomotor system, we assumed the fixation duration as a random variable with probabilistic distribution [6, 17, 18].

3.1 Marginal Distribution of Fixation Durations

Figure 3.1 shows the marginal distribution of fixation durations, which includes all of the 158,643 fixation durations of 18 participants. The x-axis represents the duration time, and the y-axis represents the log scale of the number of fixations across the whole data.

The shape of the marginal distribution of fixation durations (Figure 3.1) is roughly illustrated as an exponential function. While the distribution of the *reading* fixation durations has a quiet different shape [13], segmented into three parts, slow-rising ones for short fixations, fast-rising period until around 180 ms, and following a long tail for long fixation durations. These facts allow us to think that 180 ms of the fixation duration for reading is the most general case, but obviously not for watching the video.

Fixation durations which were longer than about 2 seconds are getting more unpredictable along with increasing the fixation duration. Though it is due to the logarithm increasing the sampling variance, occasionally the content of the video stimuli determined how long the eye gaze was fixated. In other words, the fixation more tends to maintain the gaze position when the visual constraint imposes. The visual constraints had various forms, we will discuss it in Section Long Fixation Durations.

3.2 Individual Distributions of Fixation Durations

Interpersonal differences were also examined. The individual distributions of fixation durations are shown in Figure 3.2. The medians of fixation durations from all 18 participants varies from 133 ms to 267 ms with the mean is 183.3 and the standard deviation is 41.7. For the analysis, we randomly selected 8 participants' data in Figure 3.2.
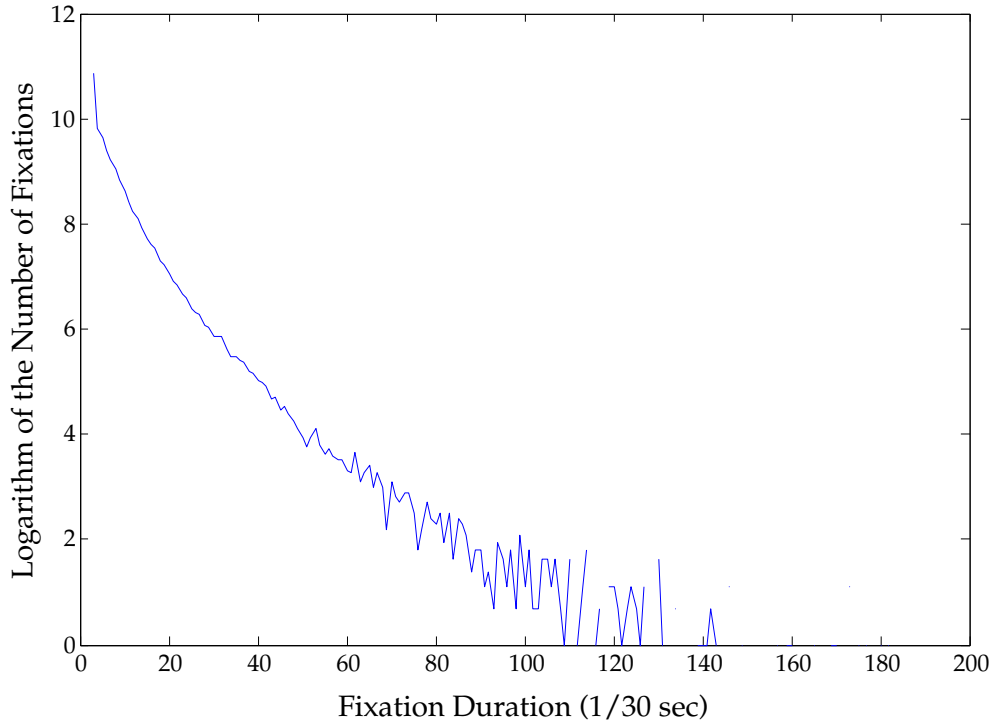
Figure 3.1 The marginal distribution of fixation durations. All of the 158,643 fixation durations of 18 participants was used. The x-axis represents the number of time unit, 1/30 sec., therefore, 30 indicates 1 second. The y-axis represents the logarithm of the number of fixation which have the same time length with regard to the time unit of x-axis.
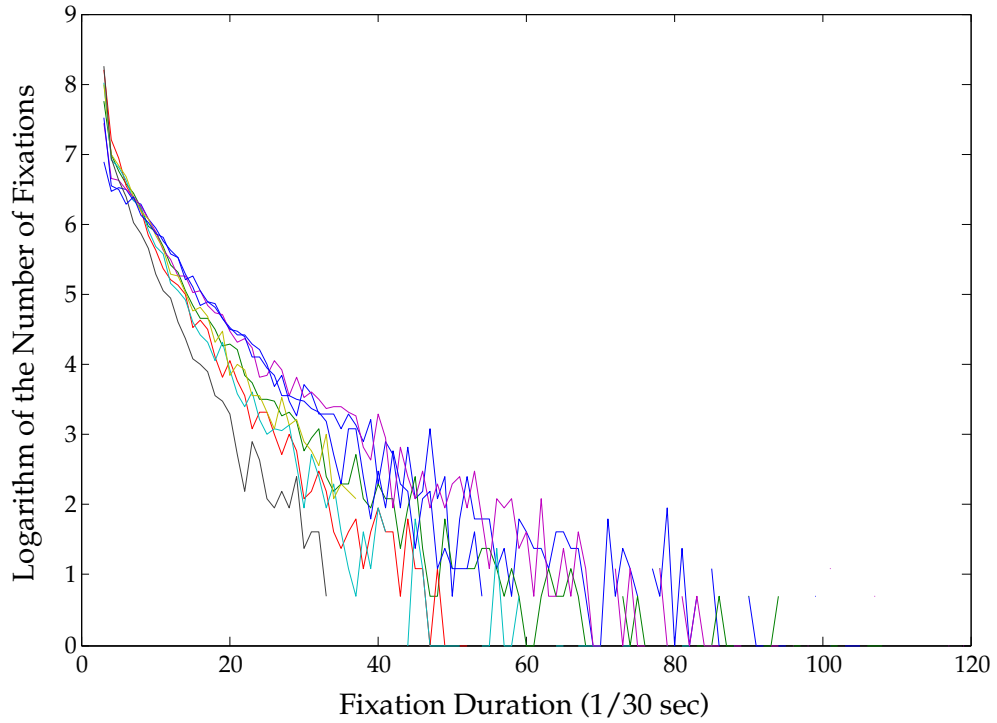
Figure 3.2 The distribution of fixation durations for the randomly selected 8 participants. We found the interpersonal differences among all 18 participants, the median of fixation durations varies from 133 ms to 267 ms (mean=183.3, std=41.7).

3.3 Long Fixation Durations

In Figure 3.3 the sequences of frames which received more than 2 seconds of the fixation duration from at least 3 different participants are shown. We set the threshold to reduce the interpersonal variation. We got all 41 sequences across 1 hour 7 minutes 50 seconds length of the material. For the review, we chose 10 typical sequences. Each row means an independent sequence and each column means a single frame. The time interval between the frames is 0.5 seconds. The colored dots mean the fixation positions, whose durations are longer than 2 seconds. The same color means the same participant. Four different types of the sequences are listed as *Alerted* (3), *Successive* (3), *Stationary* (3), and *Unclassified* (1) for the review. The classification is conducted by one of authors and two other colleagues. Every sequence used by this study is classified by agreement. If there is a conflict or not fit to the major types, then the sequence is classified as *Unclassified*.

The sequence of the *Alerted* type was classified as the scene implied an unusual and, potentially dangerous or difficult situation, which may introduce a mental arousal. The sequence of the first row demonstrates an urgent moment that a huge snowball is about to roll down back on the hill, which was previously rolled up by the robot, *Rody*. Second shows that *Pororo* has been fishing at the ice hole, but what he caught is *Shark*, a naughty character. Third shows that *Eddy* rolled his eyes to kick his ball avoiding the opponent *Pororo*.

The *Successive* type shows that the fixation duration extends across more than 2 different scenes. Because the location of the target object is not changed or changed within the range of a foveal or central vision, 2-5°, the fixation holds its position [19]. The fourth sequence shows the closed-up characters are serially shown up in the center of the screen. The fixations duration of the fifth and sixth

Figure 3.3 The sequences of frames which received more than 2 seconds of fixation duration from at least 3 participants. Each row means an independent sequence and each column means a single frame. The time interval between frames is 500 ms. The colored dots mean the fixation positions, whose durations are longer than 2 seconds. The same color of dots means what is from the same participant. First 3 sequences show the *Alerted* type, then next 3 sequences show the *Successive* type, then next 3 sequences show the *Stationary* type, and then the last sequence shows the *Unclassified* type of visual constraints for the review. For the description of those sequences, see the text in Section *Long Fixation Durations.*

sequences extends across different scenes have different visual configurations.

The *Stationary* type most clearly shows the characteristics that the indifferent scene maintains while the target object moves a little bit or even does not move. For there is not a particular event or a change of the scene, the participants tend to fixate their gazes. See the seventh through ninth sequences.

The sequence of the *Unclassified* type takes various forms. The tenth sequence shows that *Pororo* and *Crong* just jumped out of the shoulder of the magician dragon, *Tongtong* who is flying in the sky. Two sunflowers spring *Pororo* and *Crong* into the sky in multiple times. But due to the perspective of the camera, the location of the two characters in the scene is almost fixed. Participants fixated their eye gazes on the center of two characters while the background shifting up and down. In other cases, though not reported in Figure 3.3, there were cinematic techniques, i.e., tracking, tilt, zoom-in and zoom-out, and other uncertain ones. The number of these cases is relatively so small compared to the other types, hence we classify all of them into the *Unclassified*. We summarize those three typical types in Table 3.1.

Table 3.1 Long Fixation Types.

Long fixation type	Description
Alerted	An urgent situation happens with an object, which can be easily targeted as a cause.
Successive	Successive changes to keep attracting.
Stationary	The scene is the same while the target object(s) moves a little or even does not move.
Unclassified	Cinematic techniques, which are tracking, tilt, zoom-in and out, and others.

Chapter 4

Long-Term Memory Formation

In the studies of reading eye movements, as noted before, the fixation duration is a good indicator for information processing. The stimulus-response model seems to offer a way of understanding, because a sequence as a stimulus supplies a cause to response, in this case, a long fixation. Even so, we have to be cautious that the long fixation itself was not always induced when the internal state of a subject is affected. In addition, the sequences received a long fixation did not decisively guarantee the quality of information processing nor its specificity. When we carefully look into the sequences in Figure 3.3, the fixations on *Successive* or *Stationary* type sequences can be interpreted as looking passively or even blankly.

This view is also valid for the formation of memory. A stimulus is memorized by the constructive activities, a series of being stimulated, giving attention, and acquisition. Yet the corresponding responses are not deterministic by the stimulus for the uncertainty of environmental perturbations or the complexity of internal states. Therefore, the formation of memory is the result of the cogni-

tive process of response rather than the response itself. However, the cognitive process is an internal procedure, which only can be measured by the tangible responses indirectly. So, we need to discern the reliable indicator of the cognitive process with a sufficient care.

4.1 Recognition Test

We defined two types of fixation as long and short fixations. The long fixation was defined by the gaze holding its position after fixation filtering longer than 1400 ms. The short fixation was defined by the one that shorter than 300 ms. Figure 4.1 shows the memory test result for the two fixation types. Each participant assessed in a recognition test rating 8 long fixated movie clips, 8 short fixated movie clips and 4 not-seen movie clips, which were not seen previously. Contrary to our expectations, the scores for the short fixated movie clips was not much different from the the scores for the long fixated movie clips. This result makes sense when we consider that the long fixation is an attentive response, which can be actively or passively motivated by a reciprocal process in visual system.

The relationship between the emotional arousal and the formation of long-term memory was known for the studies in neuroscience [14, 15]. In this study, we defined the emotionally arousal events are urgent, threat, tension, hurt, trouble, surprise or angry situations. On top of this, more detailed analyses are conducted with regard to this.

Figure 4.3 shows The memory test result for the long fixation which was on the *Alert* movie clip or the *No Alert* movie clip. The *Alert* movie clips were including the emotional arousal events, which was previously defined. This classification is rather definite because the content is an animation video for

children. Figure 4.2 shows the result of arousal effect on the recognition test, *Alert* movie clips significantly got higher scores than *No Alert* movie clips with the p-value of 0.0077 (p < 0.01). The number of ratings from 11 participants were 19 for the *Alert* and 69 for *No Alert*. The difference between two mean scores for the long and short fixation types was significant. The p-value of two-sample t-test is 0.0104 (< 0.05). The blue horizontal line indicates the mean scores of the long fixated movie clips.

Figure 4.4 shows the memory test result for the short fixation which was on the *Alert* movie clips or the *No Alert* movie clips. The number of cases from 11 participants were 21 and 67, respectively. The p-value of two-sample t-test was 0.2484 (> 0.05). In the short fixation cases, there was no significance between two content types, *Alert* type and *No Alert* type. The blue horizontal line indicates the mean scores of the short fixated movie clips.

4.2 Gaze Variations

The difference of the eye movements between the *Alert* movie clips and the *No Alert* movie clips for the long fixated was observed by the variation of gazes. Actually, the gazes keep moving in the fixation state. Because eye balls are fixed by twitching extraocular muscles. However, the most contribution to that variation was a smooth pursuit. In the smooth pursuit, the eyes were following a slowly moving object or interesting features. Because our experimental settings did not and can not precisely define the smooth pursuit, all the eye movements slower than 30 degree per second were candidates.

To get the variation of the gazes excluding the variation from the smooth pursuit, we used the window sliding technique. Varying the size of a time window, we moved the time window on the time series by 1/30 second in every
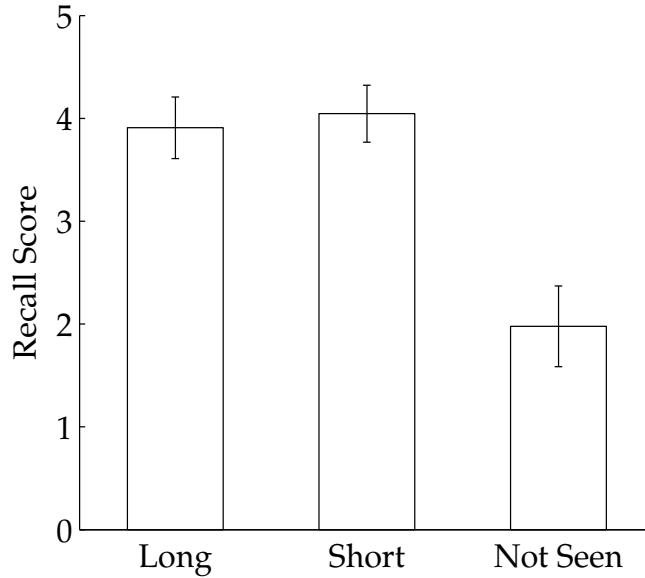
Figure 4.1 Long-term memory test result for two fixation types, one is that whose duration of fixation is longer than 1400 ms, and the other one is that whose duration of fixation is shorter than 300 ms. Each participant rates 8 long fixated movie clips, 8 short fixated movie clips and 4 control movie clips which are not seen previously. Surprisingly, the mean scores of the short fixated movie clips is indifferent to the mean scores of the long fixated movie clips (p = 0.5051). Error bars indicate ± 2 standard errors of means (SEMs). For the detail of test, refer to *Materials and Methods*.
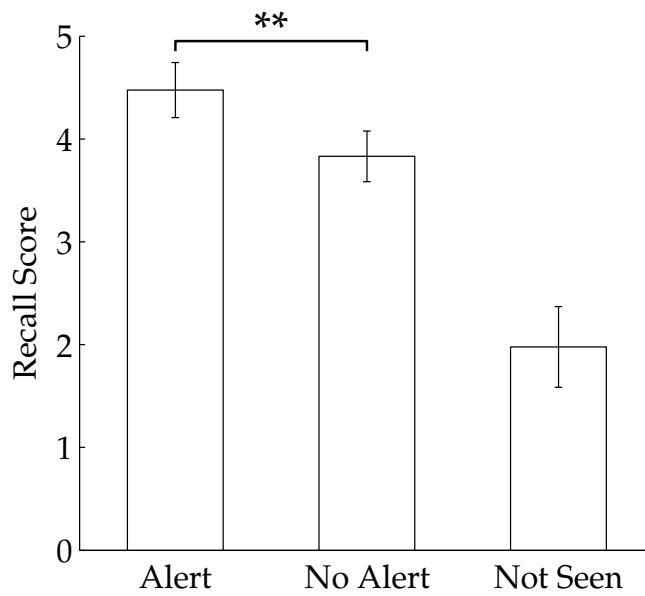
Figure 4.2 Long-term memory test result for *Alert* and *No Alert* types. The arousal affects on the recognition test, Alert movie clips significantly get higher scores than No Alert movie clips with the p-value of 0.0077 ($p < 0.01$). Error bars indicate ± 2 standard errors of means (SEMs).

Figure 4.3 The memory test result for the long fixations which are on the *Alert* movie clips or the *No Alert* movie clips. The number of cases from 11 participants are 19 and 69, respectively. The difference between two means is statistically significant, the p-value of two-sample t-test for the *Alert* type is 0.0104 (< 0.05). The blue horizontal line indicates the mean scores of the long fixated movie clips. Error bars indicate $\pm$ 2 SEMs.
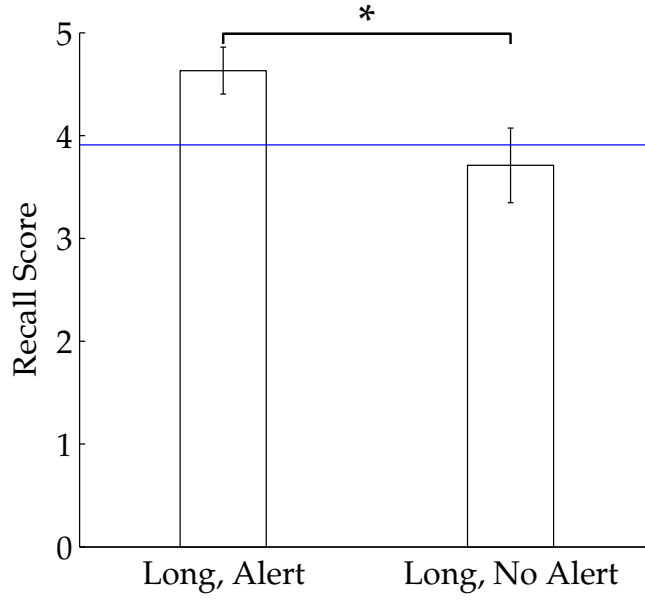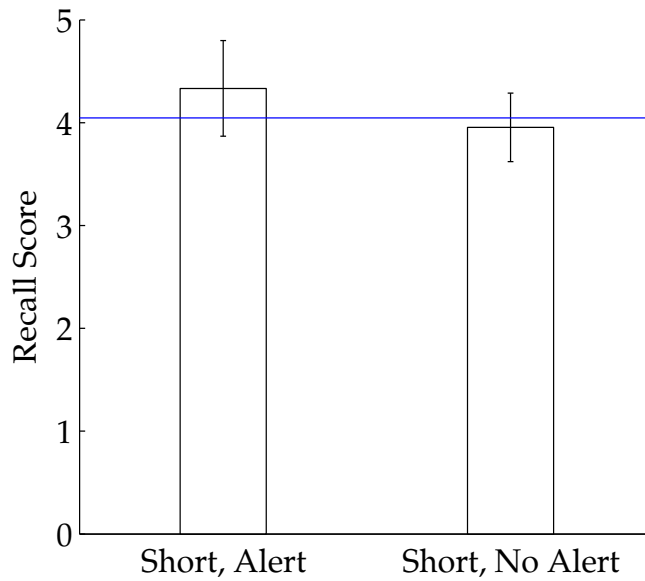
Figure 4.4 The memory test result for the short fixations which are on the *Alert* movie clips or the *No Alert* movie clips. The number of cases from 11 participants are 21 and 67, respectively. The p-value of two-sample t-test for the *Alert* type is 0.2484. The blue horizontal line indicates the mean scores of the short fixated movie clips. Error bars indicate ± 2 SEMs.

steps, and took the median of the variations. Figure 4.5 shows the significant level changes according to the change of the window size. When the window size was smaller than 200 ms, p-values were lower than 0.05. In the long fixated group, the *Alert* movie clips tended to have smaller gaze variations than the *No Alert* movie clips have, whereas the short fixated group did not show these differences.

4.3   Computational Model

Based on the study of bottom-up attention [20], the computational model was implemented by the saliency map-based approach [21]. This model used the visual information, i.e., intensities, color opponencies and orientations, as a source to measure the conspicuity [22]. While reflecting physiological evidences for the basis mechanisms of visual information processing, active selection to increase the information gain, which is depicted as "best question to ask" [23, 24], the implement had both processing efficiency and robustness to noises. It took an image as input, and it returns one saliency map, the matrix, note that, whose elements were normalized iteratively and nonlinearly [25].

Using the saliency map from the computational bottom-up attention model, we examined the association between the gaze fitness to the model and the score. The fitness function is described below. We used the SaliencyToolbox with default parameters for getting the saliency map to analyze, and aligned the gaze coordination to the area of the saliency map, which had a smaller size after sub-sampling [26].

$$X_{saliency}^{(i,j)} = \sum_{t=1}^{T} \mathcal{S}_{x_t,y_t}^{(i,j,t)} \tag{4.1}$$

In Equation 4.1, x_t and y_t represent the mapped gaze coordination at t time
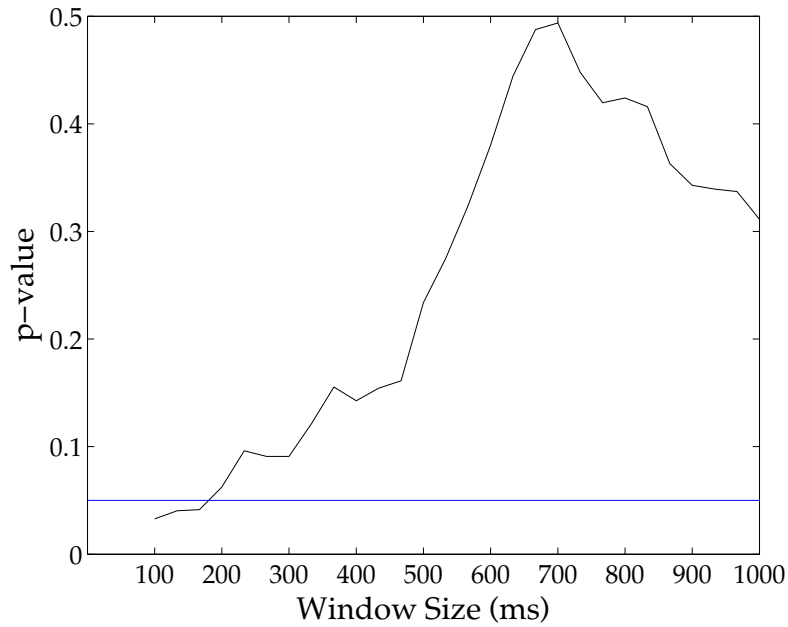
Figure 4.5 The significant levels of gaze variations with regard to the window sizes. When the window size is narrower than 200 ms, the medians of gaze variations for the *Alert* movie clips are significantly different from those for the *No Alert* movie clips ($p < 0.05$). The blue horizontal line shows the 5% significant level.

in a fixation duration. $\mathcal{S}^{(i,j,t)}$ is a saliency map which is the output of the model for a given screenshot at t time in the duration of a movie clip, which was given to the participant i for the j-th movie clip of the recognition test.

We fit the parameters for the linear regression model for the scores, which was defined by Equation 4.2,

$$Y_{score} = \mathcal{B}_0 + \sum_{l \in \mathcal{L}} \mathcal{B}_l \cdot X_l \qquad (4.2)$$

, $\mathcal{B}$s are the coefficients of the model, and $\mathcal{L}$ is a set of features as

$$\mathcal{L} = \{duration, saliency\}. \qquad (4.3)$$

$X_{duration}$ is the fixation duration in second, and $X_{saliency}$ is defined by Equation 4.1. We used only the data of the long fixations, because the short fixated movie clips did not show the significant differences on the scores for the duration and the gaze fitness to the saliency map.

Table 4.1 Estimated coefficients of the linear regression model.

Coef.	Estimate	SE	tStat	pValue
$\mathcal{B}_0$	4.8241	0.48622	9.9217	7.39e-16
$\mathcal{B}_{duration}$	-0.42214	0.18627	-2.2663	0.025974
$\mathcal{B}_{saliency}$	0.032642	0.015957	2.0455	0.043893

Table 4.1 shows the estimated coefficients of the linear regression model for the prediction. All shown parameters were statistically significant, though the gaze variation was excluded for the fitting due to its unexplainable for the scores. The gaze fitness reasonably explained the score with the p-value 0.043893. The estimated value for the gaze fitness was positive, which means

the score positively correlated with the gaze fitness in the model. Interestingly, the fixation duration, which was longer than 1400 ms, negatively correlated with the score, though the standard error of that was relatively high. The assessment of the model is shown in Table 4.2. The number of observation was 88 for each of 11 participants rates 8 long fixated movie clips, respectively. The other models, like logistic regression and non-linear regression, also examined, but they did not explain better than the linear model.

Table 4.2 Assessment of the linear regression model.

Attribute	Value
# of observations	88
Error degree of freedom	85
RMSE	1.34
R^2	0.106
Adjusted R^2	0.0847
F-statistics vs. constant model	5.02 (p-value = 0.00866)

Chapter 5

Discussions

The marginal distribution of the fixation duration shows the characteristics of the response toward the visual stimuli. The shape of the marginal distribution of fixation duration can be estimated as the exponential function though the marginal distribution of the reading fixation durations is illustrated as a left-skewed normal distribution, peaking at 180 ms. Then why reading and watching are so different from each other in this property? The first to think is the difference of the cognitive process during the fixation. Simply put, reading involves the visual processing in addition to the lexical processing. The visual process captures the letters through the retina, Lateral Geniculate Nucleus (LGN) and the primary visual cortex. Then the information of the letters is directed to the distributed lexical processing areas. Though skipping is occasionally occurred while reading, those serial processes spend a bit of latency time. Since reading is an active task, the decision that when to move and where to move to a next word, is actively and consciously made comparing to watching task, those latencies tend to be preserved the normative shape. By the way, watching the

video has a different condition. The lexical processes are typically not needed, which are heavy tasks in time. Also the video stimuli are passive in regard to the temporal aspect. Therefore the generation of the long fixation duration is sufficiently constrained by the duration of the stimulus, at the same time, the content of the stimulus.

We recap the visual constraints which trigger the long fixation durations as three types in Table 3.1. The *Alerted* is an ongoing urgent situation that makes the eyes fixates an object which is thought to be a cause or a factor. This type potentially induces the emotional arousal. The second type is the *Successive*. It looks like the successive appearing of the objects keeps the fixation longer, however, the interpretation that the successive absence of the other attracting elements just lets the fixation persist is also possible. The *Stationary* shows indifferent movie clips and there is no significant change. Many movie clips show calm and relaxed or depressed situation. Emotionally, the opposite of *Alerted*. After all, what is the meaning of the long fixation durations on the video stimuli? Despite of the fact that it could be a latency time to process the cognitive information of the visual stimuli, in the other perspective, it is an waiting time to the potentially salient moment on that eye position. The waiting on the prospective location for a dramatic change or a new event is an efficient way of information processing.

Though the long fixation itself does not describe the presence of the cognitive process to memorize, in the condition of the long fixation, the arousal effect is remarkable than the others in the condition of the short fixation. As we discuss before, the long fixation is induced by active taking or passive exposure. For the cause of the long fixation is relatively well established by those two factors, the arousal effect on the long fixation is noticeable, but for the cause of the short fixation is complex and vague, the arousal effect on the short fixation is

not observable.

The characteristics of the eye movement on the arousal effect are probed by the statistical method and the computational attention model. First of all, we hypothesize on that the variation of gaze points indicates the observable response of the arousal effect in the condition of the long fixation. As we report in Section Gaze Variations, the gaze variation on the alerted movie clips are smaller than those on the other case, when the window size is shorter than 200 ms, to minimize the variation from the slow pursuit. Yet, the gaze variation does not have a statistical power to predict on the score, which is the measurement for the long-term memory. Though the arousal stimuli associate with the long-term memory [14, 15], we conclude that the gaze variation explains only the arousal effect, not until the further cognitive process.

Second, we establish the linear model to predict the score using the computational attention model [21] in addition to the fixation duration. For the short fixated movie clips, we cannot find the predicting variables, so we limit to use the observed data of the long fixated movie clips. This is the backward study of Itti (2006)'s work [27], which investigates the fitness of the model to human gaze, whereas our work uses the model to evaluate the attentiveness of human. In machine learning, Zou et al. (2012) reported that stimulated fiaxations help to capture the useful invariant features in the image recognition task [28]. This work gives a hint from a computational approach for the account.

Although we formulate the simple summation of saliency scores, it shows the significant level for the prediction. The representation of saliency map is also found in the neuronal structures [29], which have a peak avitivity in the physically salient position. And, if the fixation duration is longer than 1400 ms, we observe that the longer fixate, the less memorize. Nevertheless, we have to be cautious enoughly for the standard error of this variable is a bit of high.

The value of the adjusted R^2 for the model reminds us that the information of eye movement should carefully use to estimate the cognitive process. As well as, there is the limitation of the analysis coverage, for the majority of the eye movement is the short fixated ones.

However, the investigation is not complete. We look forward to have more distinct features. A temporal modeling using the salient detectors [30, 31] and the optic flow [32] may be a promising option. Moreover, the cognitive modeling for the scene comprehension, which is related to the emotion-based reaction, guides us into the diffent level of a methodological stage.

As we discussed, the long fixations are constrained by the visual content of the video stimuli, and with regard to that, the eye movements are planned in a reciprocal manner [33]. The estimation of the eye movement is viable on top of a reciprocally anticipatory model [34]. The serial information of the fixation positions can be used to parsimoniously select the portion of the visual features on the scene for an application using the features, like a human, it can be a breakthrough for the cognitive modeling on the endless stream of the visual information.

Chapter 6

Conclusions

We study the characteristics of the eye movements through the marginal distribution of fixation durations. We notice that the marginal distribution of fixation durations for the video stimuli has a form which is different from the marginal distribution of fixation durations for the reading materials. The behavioral basis for the difference may attribute to the lighter load for the cognitive process and the temporal and spatial constraints which are given by the video stimuli. Those constraints are summed up as three distinctive types, Alerted, Successive and Stationary. We find the arousal effect only for the long fixated movie clips. The small gaze variation for the long fixation indicates an active response to the arousal stimuli. However, the gaze variation does not help to estimate for the score, among the long fixations, just the fixation duration and the saliency-probing activity are significant to model. We can appreciate this computational model within the embodied cognitive framework with the perception-action cycling.

Bibliography

[1] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2.   IEEE, 2005, pp. 1800–1807.

[2] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2.   IEEE, 2006, pp. 2169–2178.

[3] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 2411–2418.

[4] J. Henderson, "Human gaze control during real-world scene perception," *Trends in Cognitive Sciences*, vol. 7, no. 11, pp. 498–504, 2003.

[5] M. Zirnsak and T. Moore, "Saccades and shifting receptive fields: anticipating consequences or selecting targets?" *Trends in Cognitive Sciences*, Nov. 2014.

[6] K. Rayner, "Eye Movements in Reading and Information Processing: 20 Years of Research," *Psychological Bulletin*, vol. 124, no. 3, pp. 372–422, 1998.

[7] E. D. Reichle, A. Pollatsek, D. L. Fisher, and K. Rayner, "Toward a Model of Eye Movement Control in Reading," *Psychological Review*, vol. 105, no. 1, pp. 125–157, 1998.

[8] K. Rayner, "Visual attention in reading: Eye movements reflect cognitive processes," *Memory & Cognition*, vol. 5, no. 4, pp. 443–448, 1977.

[9] A. W. Inhoff and K. Rayner, "Parafoveal word processing during eye fixations in reading: Effects of word frequency," *Perception & Psychophysics*, vol. 40, no. 6, pp. 431–439, 1986.

[10] K. Rayner and S. A. Duffy, "Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity," *Memory & Cognition*, vol. 14, no. 3, pp. 191–201, 1986.

[11] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: reinterpreting salience." *Journal of vision*, vol. 11, no. 5, p. 5, 2011.

[12] J. M. Findlay and R. Walker, "A model of saccade generation based on parallel processing and competitive inhibition." *The Behavioral and brain sciences*, vol. 22, no. 4, pp. 661–674; discussion 674–721, 1999.

[13] G. Feng, "Eye movements as time-series random variables: A stochastic model of eye movement control in reading," *Cognitive Systems Research*, vol. 7, no. 1, pp. 70–95, 2006.

[14] L. Cahill, R. J. Haier, J. Fallon, M. T. Alkire, C. Tang, D. Keator, J. Wu, and J. L. McGaugh, "Amygdala activity at encoding correlated with long-term, free recall of emotional information," *Proceedings of the National Academy of Sciences*, vol. 93, no. 15, pp. 8016–8021, 1996.

[15] L. Cahill and J. L. McGaugh, "Mechanisms of emotional arousal and lasting declarative memory," *Trends in Neurosciences*, vol. 21, no. 7, pp. 294 – 299, 1998.

[16] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*.   New York, New York, USA: ACM Press, 2000, pp. 71–78.

[17] E. D. Reichle, K. Rayner, and A. Pollatsek, "The E-Z Reader model of eye-movement control in reading: Comparisons to other models," *Behavioral and Brain Sciences*, vol. 26, no. 04, pp. 445–476, 2004.

[18] E. D. Reichle, A. Pollatsek, and K. Rayner, "E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading," *Cognitive Systems Research*, vol. 7, no. 1, pp. 4–22, 2006.

[19] T. McMorris, *Acquisition and performance of sports skills*.   John Wiley & Sons, 2014.

[20] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry." *Human neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[21] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[22] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision Research*, vol. 42, no. 1, pp. 107–123, Jan. 2002.

[23] P. Reinagel and A. Zador, "Natural scene statistics at the centre of gaze," *Network: Computation in Neural Systems*, vol. 10, no. 4, pp. 341–350, Nov. 1999.

[24] C. Zetzsche, K. Schill, H. Deubel, G. Krieger, E. Umkehrer, and S. Beinlich, "Investigation of a sensorimotor system for saccadic scene analysis: an integrated approach," in *Proc. 5th Int. Conf. Simulation Adaptive Behav*, vol. 5, 1998, pp. 120–126.

[25] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision research*, vol. 40, no. 10, pp. 1489–1506, 2000.

[26] D. Walther and C. Koch, "Modeling attention to salient proto-objects." *Neural networks : the official journal of the International Neural Network Society*, vol. 19, no. 9, pp. 1395–407, Nov. 2006.

[27] L. Itti, "Quantitative modelling of perceptual salience at human eye position," *Visual Cognition*, vol. 14, no. 4-8, pp. 959–984, Aug. 2006.

[28] W. Y. Zou, S. Zhu, A. Y. Ng, and K. Yu, "Deep learning of invariant features via simulated fixations in video," *Advances in Neural Information Processing Systems*, pp. 3212–3220, 2012.

[29] J. H. Fecteau and D. P. Munoz, "Salience, relevance, and firing: a priority map for target selection," *Trends in Cognitive Sciences*, vol. 10, no. 8, pp. 382–390, Nov. 2006.

[30] D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.

[31] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.

[32] J. J. Koenderink, "Optic flow," *Vision research*, vol. 26, no. 1, pp. 161–179, 1986.

[33] B.-T. Zhang, "Information-theoretic objective functions for lifelong learning." in *AAAI Spring Symposium: Lifelong Machine Learning*, 2013.

[34] R. Robert, "Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations," *Pergamon Press*, 1985.

초록

안구 운동은 뇌 활동을 관찰할 수 있는 비침습적이면서 활용도가 높은 지표이다. 이러한 배경으로 최근 안경류 웨어러블 장비가 연구 기관을 중심으로 빠르게 보급되고 있으며 안구 운동에 대한 연구 활동과 관련 산업의 성장을 고도화 하고 있다. 우리는 아동용 비디오 시청 환경에서 안구 운동을 장기 응시와 단기 응시를 분류하여 연구해 보았다. 응시 기간에 따른 장기 기억 상관 여부를 확인해 보았지만 유의성을 발견할 수 없었다. 하지만 시청 자극을 주의 상황과 중립 상황으로 나눈 후 응시 기간에 따른 차이를 확인 하였을 때 긴 응시에서는 장기 기억 인출에서 통계적으로 유의미한 차이를 발견할 수 있었지만 짧은 응시에서는 그렇지 못하였다. 이를 바탕으로 관측된 안구 운동에 대한 계산학적인 방법인 현저점 탐색과 응시 길이를 이용한 선형 회귀 모델을 제안하고 평가하였다. 이러한 안구 운동의 특징에서 발견되는 장기 기억과 관련된 인지 처리 과정은 평생 학습 기제에서 선택적 주의를 통한 효율성 추구가 제한된 인지 처리 능력 안에서 중요한 요소로 기여함을 시사한다.

Acknowledgements