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

Jin-Hwa Kim (jhkim@bi.snu.ac.kr)
Cognitive Science Program, Seoul National University
Seoul, 151-742, Republic of Korea

Byoung-Tak Zhang (btzhang@bi.snu.ac.kr)
School of Computer Science and Engineering, Seoul National University
Seoul, 151-742, Republic of Korea

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 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 sequences statistically have higher recall scores, but the short fixations do not. Finally, we tentatively propose a computational model for the embodied cognitive framework with the perception-action cycling, which may provide an explanatory way to the efficient memory mechanism for the life-long sequences.

Keywords: Eye movement; fixation; spatio-temporal; long-term memory, computational modeling.

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 (Winn, Criminisi, & Minka, 2005; Lazebnik, Schmid, & Ponce, 2006), and object tracking (Wu, Lim, & Yang, 2013). 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 understand 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 more complicated how the brain works for the acquisition of the visual information (Henderson, 2003). 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.

The studies on the reading eye movement, which are relatively well studied by psychologists and neuroscientists (Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998), reveal the fixation duration is related to the presence of the cognitive process (Rayner, 1977), for an instance, observing its correlation with linguistic attributes (Inhoff & Rayner, 1986; Rayner & Duffy, 1986).

There is a different aspect of studying on the eye movement for video stimuli. How long the fixation sustains is more constrained by the affective content, i.e. emotional response, context of the content, rather than the reading materials do. And 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 or avoided due to the research complexities and the methodological difficulties (Tatler, Hayhoe, Land, & Ballard, 2011). However, the recent advancement of the sensory device, like the Google Glass and the mobile devices for the eye tracking, promotes to study on the video stimuli and even more natural experimental environment, and to implement the research model or applications on those mobile devices.

We investigate the characteristics of eye movements toward the video stimuli. 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 (Findlay & Walker, 1999; Feng, 2006). In this study, we focus on the characteristics of the fixation duration as the evidence of the cognitive process. Moreover, as the sequences which potentially induce the emotional arousal are known for helping to recall the seen movie clips (Cahill et al., 1996; Cahill & McGaugh, 1998), we will see if the arousal effect is asserted by the duration of fixation.

Materials and Methods

Experiment 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 who have marked individualities. *Pororo Season 3 DVD 1* contains the 13 consecutive episodes that each has a single storyline.

We recruit 18 participants with normal vision (11 males, 6 females; 23-31 years of age; median is 25), who are voluntarily participate in the study. All participants had not experienced a brain damage or a behavioral disorder. The participants are new to the video, *Pororo Season 3*. Experimental procedures were agreed by the Research Ethics of Seoul National University and the participants signed a consent form.

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

Concurrently, the eye movements and the user-perspective scenes are 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 classifies fixations with the velocity threshold, 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 reasonable approach (Salvucci & Goldberg, 2000).

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

Experiment 2

For 11 participants who have participated in *Experiment 1*, we prepared the controlled memory test for each participant. We conducted this experiment 3-4 months later after *Experiment 1* (the intervals are not consistent due to schedule conflicts). A memory test consists of total 20 video sequences; 8 sequences for long fixations, another 8 sequences for short fixations, and the remaining 4 sequences for the control, which are not seen in the previous experiment. In detail, the lengths of all video sequences are the same as 3 seconds. *The long fix-*
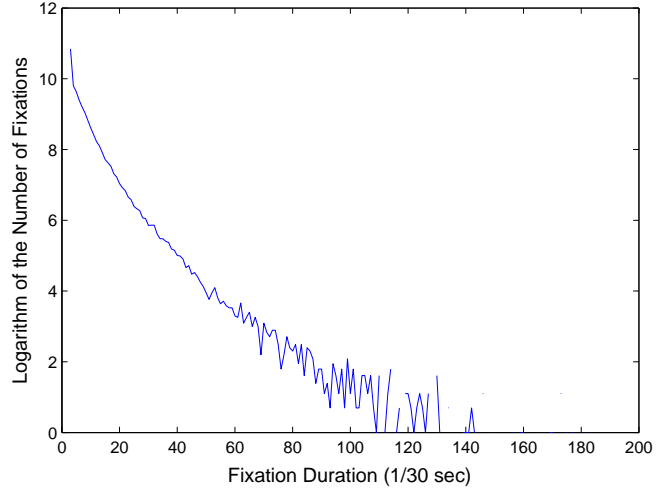

Figure 1: The marginal distribution of fixation durations. All of the 158,643 fixation durations of 18 participants is used. The x-axis represents the number of time unit, 1/30 sec., therefore, 30 indicates 1 second. The y-axis represents the logarithm of the number of fixation which have the same time length with regard to the time unit of x-axis.
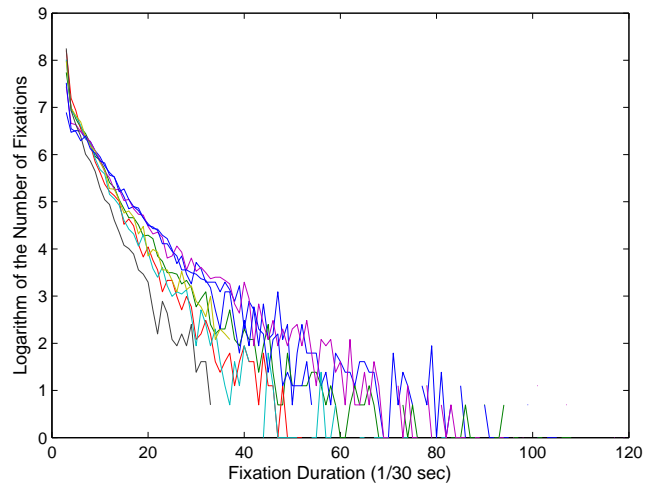

Figure 2: The distribution of fixation durations for the randomly selected 8 participants. We find the interpersonal differences among all 18 participants, the median of fixation durations varies from 133 ms to 267 ms (mean=183.3, std=41.7).

ation sequences are randomly picked from each participant's data containing the fixation longer than 1400 ms in the middle of the sequence. *The short fixation sequences* are randomly picked from each participant's data containing the fixation shorter than 300 ms. The 4 control sequences are randomly picked from the other season of *Pororo* series, *Pororo Season 2*.

Each participant identifies 20 sequences randomly sorted. Each participant gives a sequence 1 to 5 score depending on the assurance of whether he or she saw the sequence before or not.

Fixation Duration

Fixation duration takes a important position in the reading literature for the durations of eye fixations seems to be constrained by the linguistic features of the fixated word (Rayner & Duffy, 1986; Inhoff & Rayner, 1986). In the video watching task, we anticipate the characteristics of fixation duration are different from those in the reading task. As we expect, the fixation duration changes more drastically, up to 10 seconds during watching the video. These changes are 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 the random variable with probabilistic distribution (Rayner, 1998; Reichle, Rayner, & Pollatsek, 2004; Reichle, Pollatsek, & Rayner, 2006).

Marginal Distribution of Fixation Durations

Figure 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 1) is roughly illustrated as an exponential function. While the distribution of the *reading* fixation durations has a quiet different shape (Feng, 2006), 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 are 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 determines how long the eye gaze is fixated. In other words, the fixation more tends to maintain the gaze position when the visual constraint imposes. The visual constraints have various forms, we will discuss it in subsection Long Fixation Durations.
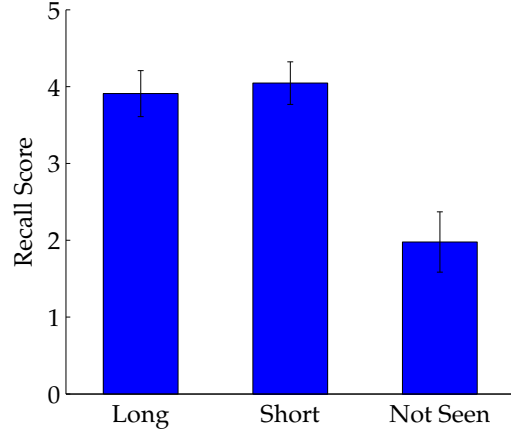

Figure 3: Long-term memory test result for two fixation types, one is that whose duration of fixation is longer than 1400 ms, and the other one is that whose duration of fixation is shorter than 300 ms. Each participant rates 8 long fixated sequences, 8 short fixated sequences and 4 control sequences which are not seen previously. Surprisingly, the mean scores of the short fixated sequences is indifferent to the mean scores of the long fixated sequences (p = 0.5051). Error bars indicate $\pm$ 2 standard errors of means (SEMs). For the detail of test, refer to *Materials and Methods*.

Individual Distributions of Fixation Durations

Interpersonal differences are also examined. The individual distributions of fixation durations are shown in Figure 2. For the analysis, it only shows randomly selected 8 participants' data. 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.

Long Fixation Durations

In Figure 4 the sequences of frames which receive 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 sequence of the *Alerted* type is classified as the scene implies 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

Figure 4: The sequences of frames which receive more than 2 seconds of the fixation duration from at least 3 participants. Each row means an independent sequence and each column means a single frame. The time interval between the frames is 500 ms. The colored dots mean the fixation positions, whose durations are longer than 2 seconds. The same color means that which 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 subsection *Long Fixation Durations*.

Table 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.

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, $2\text{-}5°$, the fixation holds its position (McMorris, 2014). 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 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 4, there are 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 classified all of them into the *Unclassified*.

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 the way of understanding, because a sequence as a stimulus supplies a cause to response, in this case, a long fixation. But we have to be cautious that the long fixation itself is not always induced when the internal state of a subject is positively affected. In addition, the sequences received a long fixation does not decisively guarantee the quality of information processing nor its specificity. When we carefully look into the sequences in Figure 4, 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. But the
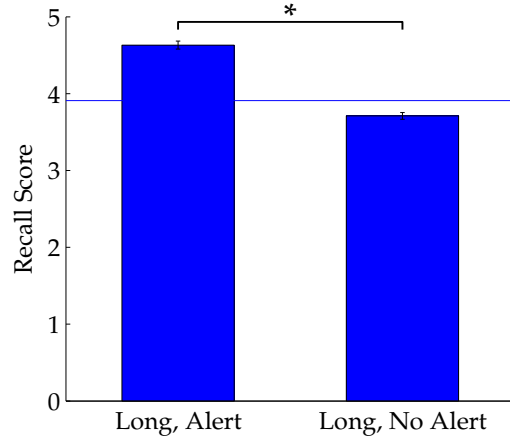

Figure 5: The memory test result for the long fixation which is on the *Alert* sequence or the *No Alert* sequence. The number of cases from 11 participants are 19 and 69, respectively. The difference between two means is statistically significant, the p-value of two-sample t-test for the *Alert* type is 0.0104 (< 0.05). The blue horizontal line indicates the mean scores of the long fixated sequences. Error bars indicate ± 2 SEMs.
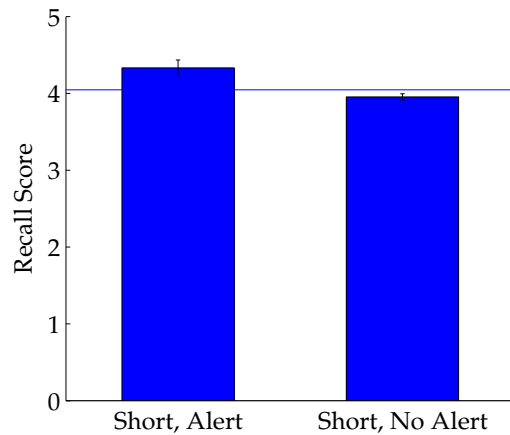

Figure 6: The memory test result for the short fixation types which is on the *Alert* sequence or the *No Alert* sequence. The number of cases from 11 participants are 21 and 67, respectively. The p-value of two-sample t-test for the *Alert* type is 0.2484 (> 0.05). The blue horizontal line indicates the mean scores of the short fixated sequences. Error bars indicate ± 2 SEMs.

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 cognitive 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.

Recall Test

We define two types of fixation as long and short fixations. The long fixation is the gaze holding its position after fixation filtering longer than 1400 ms. The short fixation is one that shorter than 300 ms. Figure 3 shows the memory test result for the two fixation types. Each participant assessed in a recall test rating 8 long fixated sequences, 8 short fixated sequences and 4 not-seen sequences, which are not seen previously. Contrary to our expectations, the recall scores for the short fixated sequences is not much different from the the the recall scores for the long fixated sequences. This result makes sense when we consider that the long fixation is an attentive response, which is actively or passively motivated by a reciprocal process in visual system.

The relationship between the emotional arousal and the formation of long-term memory is known for the studies in neuroscience (Cahill et al., 1996; Cahill & McGaugh, 1998). In this study, we define 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 the type of stimuli.

Figure 5 shows The memory test result for the long fixation which is on the *Alert* sequence or the *No Alert* sequence. The *Alert* sequences are classified by the predefined conditions, an unusual and, potentially dangerous or difficult situation, which may introduce an emotional arousal. This classification is rather definite because the content is an animation video for children. We do not include the result of arousal effect on the recall test, *Alert* sequences significantly get higher recall scores than *No Alert* sequences with the p-value of 0.0077 ($p < 0.01$). The number of ratings from 11 participants are 19 for the *Alert* and 69 for *No Alert*. The difference between two mean scores for the long and short fixation types is 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 sequences.

Figure 6 shows the memory test result for the short fixation types which is on the *Alert* sequences or the *No Alert* sequences. The number of cases from 11 participants are 21 and 67, respectively. The p-value of two-sample t-test is 0.2484 (> 0.05). In the short fixation cases, there is no significance between two content types, *Alert* type and *No Alert* type. The blue horizontal line indicates the mean scores of the short fixated sequences.
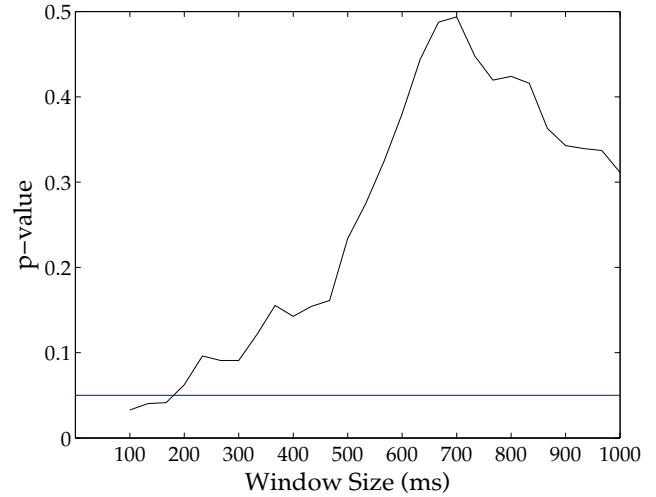

Figure 7: The significant levels with regard to the window sizes. When the window size is narrower than 200 ms, the medians of gaze variations for the *Alert* sequences are significantly different from those for the *No Alert* sequences ($p < 0.05$).

Gaze Variations

The difference of the eye movements between the *Alert* sequences and the *No Alert* sequences for the long fixated is 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 is a smooth pursuit. In the smooth pursuit, the eyes are following a moving object or interesting features. Because our experimental settings do not and can not precisely define the smooth pursuit, all the eye movements slower than 30 deg/sec are candidates.

To get the variation of the gazes excluding the variation from the smooth pursuit, we can use the window sliding technique. Varying the size of a time window, we move the time window on the time series by 1/30 second in every steps, and take the median of the variations. Figure 7 shows the significant level changes according to the change of the window size. When the window size is smaller than 200 ms, p-values are lower than 0.05. In the long fixated group, the *Alert* sequences tend to have smaller gaze variations than the *No Alert* sequences have, whereas the short fixated group does not show this difference.

Computational Model

Based on the study of bottom-up attention (Koch & Ullman, 1985), the computational model is implemented by the saliency map-based approach (Itti, Koch, & Niebur, 1998). This model uses the visual information, i.e., intensities, color opponencies and orientations, as a source to measure a conspicuity. While reflecting some physiological evidence for the basis mechanism of visual information processing, it has processing efficiency and robustness to noises. It takes an im-

Table 2: Estimated coefficients of the linear regression model for the prediction of the recall score.

Coefficient	Estimate	SE	tStat	pValue
$\mathcal{B}_0$	4.8241	0.48622	9.9217	7.3904e-16
$\mathcal{B}_{duration}$	-0.42214	0.18627	-2.2663	0.025974
$\mathcal{B}_{saliency}$	0.032642	0.015957	2.0455	0.043893

age as input, and it returns one saliency map, the matrix, note that, whose elements are normalized iteratively and nonlinearly (Itti & Koch, 2000).

Using the saliency map from the computational bottom-up attention model, we examine the association between the gaze fitness to the model and the recall score. The fitness function is described below. We use the SaliencyToolbox with default parameters for getting the saliency map to analyze, and align the gaze coordination to the area of the saliency map, which has a smaller size after sub-sampling (Walther & Koch, 2006).

$$X_{saliency}^{(i,j)} = \sum_{t=1}^{T} S_{x_t, y_t}^{(i,j,t)} \qquad (1)$$

In Equation 1, x_t and y_t represent the mapped gaze coordination at t time in a fixation duration. $S^{(i,j,t)}$ 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 sequence of the recall test.

Table 3: 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)

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 some latency time. Since reading is an active task, the subject decides when and where to move to a next word, those latencies tend to be preserved. 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 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 sequences and there is no significant change. Many sequences 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.

These various types of the long fixated sequences show that a naive approach to the predict the visual constraint which induces the long fixation using a single analytical method, like the salient feature detectors (Marr & Hildreth, 1980; Canny, 1986) or the optic flow (Koenderink, 1986) would fail. The *Successive* and *Stationary* types can be estimated by using the combination of the appropriate feature detectors and the optic flow, and the temporal modeling of these. However, the *Alerted* cannot be solved without the cognitive modeling for the scene comprehension, which is related to the emotion-based reaction.

The computational modeling of the long fixation durations is worth to discuss. As we discussed, the long fixations are constrained by the visual content of the video stimuli (Zhang, 2013), the probabilistic modeling of that would be unsuccessful to compatible with the unprecedented stimuli. The first possible attempt is the composition of the salient de-

tectors (Marr & Hildreth, 1980; Canny, 1986) and the optic flow (Koenderink, 1986) accompanying the tentative cognitive modeling of that. As there are the types of the long fixation durations, the cognitive modeling can be projected to the lower dimension of the search space. Additionally, the estimation of the eye tracking is viable on top of the model. The serial information of the fixation positions enables us to intelligently select the portion of the visual features on the scene, like a human, it can be a breakthrough for the cognitive modeling on the endless stream of the visual information.

Conclusions

We studied 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. Also, we present the tentative suggestion for the computational modeling of the eye movements using the combination of the feature detectors and the optic flow in aid of the cognitive modeling.

Acknowledgments

References

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., et al. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences*, *93*(15), 8016-8021.

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*(7), 294 - 299.

Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(6), 679–698.

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 70–95.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *The Behavioral and brain sciences*, *22*(4), 661–674; discussion 674–721.

Henderson, J. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10), 1489–1506.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(11), 1254–1259.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, *4*(4), 219–227.

Koenderink, J. J. (1986). Optic flow. *Vision research*, *26*(1), 161–179.

Lazebnik, S., Schmid, C., & Ponce, J. (2006). 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, pp. 2169–2178).

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*(1167), 187–217.

McMorris, T. (2014). *Acquisition and performance of sports skills*. John Wiley & Sons.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition*, *5*(4), 443-448.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a Model of Eye Movement Control in Reading. *Psychological Review*, *105*(1), 125–157.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). EZ Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4–22.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2004). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(04), 445–476.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on eye tracking research & applications - etra '00* (pp. 71–78). New York, New York, USA: ACM Press.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of vision*, *11*(5), 5.

Walther, D., & Koch, C. (2006, November). Modeling atten-

tion to salient proto-objects. *Neural networks : the official journal of the International Neural Network Society*, *19*(9), 1395–407.

Winn, J., Criminisi, A., & Minka, T. (2005). Object categorization by learned universal visual dictionary. In *Computer vision, 2005. iccv 2005. tenth ieee international conference on* (Vol. 2, pp. 1800–1807).

Wu, Y., Lim, J., & Yang, M.-H. (2013). Online object tracking: A benchmark. In *Computer vision and pattern recognition (cvpr), 2013 ieee conference on* (p. 2411-2418).

Zhang, B.-T. (2013). Information-theoretic objective functions for lifelong learning. In *Aaai spring symposium: Lifelong machine learning.*