# Chronological Prediction of Certainty in Recall Tests using Markov Models of Eye Movements

Noaya Takahashi and Minoru Nakayama
*Human System Science, The Graduate School of Decision Science & Technology*
*Tokyo Institute of Technology*
*Meguro, Tokyo 152-8552 Japan*
*Email: ntakahashi@nk.cradle.titech.ac.jp, nakayama@cradle.titech.ac.jp*

***Abstract*—To predict chronologically certainty levels of understanding of question statements during recall tests using observed eye movements, the hidden Markov model was employed. The feasibility of predicting the accuracy of responses was examined using optimization and simulation of the model together with experimental eye movement data. Chronological prediction accuracy could be calculated using this model, and the accuracy decreased with task difficulty. The highest accuracies were observed at 550 msec. after onset of stimuli. The certainty of correct responses was calculated using the probability of transition. This certainty was the highest during the 100-250 msec. after stimuli onset, and decreased with the duration of the response. These results provide the possibility of estimating the progress of understanding.**

***Keywords*-Eye Movements; Hidden Markov Model; Certainty; User Intention**

## I. Introduction

Most web sites ask viewers to provide responses to questions about whether they understand or are interested in the content they are viewing, such as pressing the “Like” button. Even in the human-computer interaction environment, various systems ask users if they understand the specific situation or prefer the service the system provides. For example, many web sites often invite users to make responses such as “yes”, “no” or “Like”, to confirm their understanding of the context of an idea or explanation.

According to the item response theory for tests, response correctness or validity is related to the certainty of the response [1]. Using a different approach, users can make correct responses or provide satisfactory reactions acceptable to the application software, when the level of certainty is high. The theory uses eye movements which occur during the reading of statements to estimate certainty or levels of understanding of statements [2], [3]. It has been proposed that the user’s intent can be predicted using eye fixation behavior [4]. Recently, many studies deal with predicting the viewer’s intent using eye movement features and machine learning procedures, and have achieved a high level of accuracy [5].

These approaches use all of the data from onset of stimulus to response, and the progress of decision making is not a concern. To investigate the decision making process as the reason why users accept the statements or services offered, a dynamic or time series analysis is required. The process of reading questions has been studied using eye movements to analyze the ways of understanding information which is read [6]. When some mental states are defined, transitions of states can be generated in time series, using behavioral data. For examples, the transition of cognitive states of drivers of automobiles was analyzed using the hidden Markov model (HMM) [7] using the pattern of eye movements [8]. The transition between states, which occurs when reading documents to determine their relevancy was established using HMM and features of eye movements [9]. According to previous studies, it may be possible to create a model to predict the level of certainty and the accuracy of responses using HMM while eye movements during the reading of question statements are observed. As mentioned above, the model may provide the possibility of interpolation for decision making. Analysis of the transition of states using eye movements during the early stages of reading statements will reveal the visual information process. This information may contribute to the improvement of usability of human interfaces and other HCI issues.

In this paper, the hidden Markov model was used to determine the levels of certainty of understanding by observing eye movements. The novelty of this approach is to present a possibility of conducting a dynamic or time series analysis to investigate the decision making process. The feasibility of predicting the accuracy of responses and estimating the levels of certainty were examined while the model was applied to recall tests [3].

Section 2 will summarize the related works. Section 3 will describe the experimental procedure and measurements. Section 4 will explain the development procedure of hidden Markov model for this experiment. Section 5 and 6 will evaluate and discuss the results. Section 7 will summarize our findings.

## II. Related work

As mentioned in the introduction section, many studies about assessing usability [10] and predicting user intent [4] using eye movements exist. The ways of approaching the


Figure 1. The screen on the left (a) shows a sample of a definition statement: “A theater is located on the east side of the police station.” The screen on the right (b) shows a sample of a question statement: “There is a post office on the south side of the theater.”

matter can be divided into two categories, discrimination analysis, which uses all of the data, and state transition, which uses time series data. Mental states are defined as explicit in the former and implicit in the latter.

A previous study used two interface operations (zooming-in and zooming-out), which indicate a user’s intent. Intent was predicted using a liner discrimination function with eye gaze data [4]. As machine learning studies advance, smaller sizes and sparser data sets can be used to create classifiers, though more specific features of eye movements are required. Various features which are related to a mental status, which is explicit may contribute to performance, as the extraction of features is a key issue in the development of the model. The appropriate combination of features can be used to perform discriminate analysis with a high level of accuracy, therefore the selection of these features is often discussed in detail [3], [5].

Another approach is the analysis of eye movement data as sequential data. The conventional approach is to employ scanpaths of eye movements [11]. Both fixations and saccades of eye movements are analyzed as temporal and spatial behavior data. However, analysis of scanpaths of eye movements is not useful sometimes. Markovian analysis of eye movements has been introduced to estimate user intent. In particular, HMM is preferred for the analysis of sequential data, such as speech recognition, hand writing recognition, analysis of biological sequence and others [7]. Liu (1999) defined some implicit cognitive states for drivers of cars, where the model parameters for left-to-right HMM were estimated from a data set of eye movements. This approach enabled recognition of driver’s intentions from the eye movement data [8]. HMM was also applied to an information retrieval task, which consisted of two implicit states such as relevance and irrelevance while eye movements were observed [9]. As a result, the model could provide implicit relevance feedback by making inferences from the eye movement data.

According to the previous studies, implicit states may be flexibly defined if they are related to eye movement behavior. There are a variety of possible applications. Also, sequential

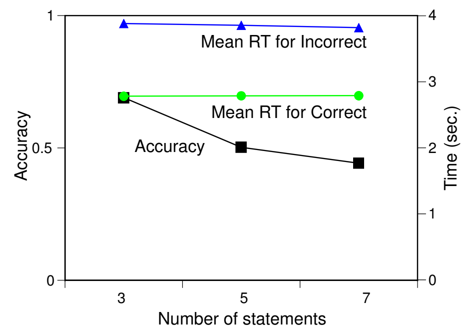


Figure 2. Correct rate and mean reaction times for responses

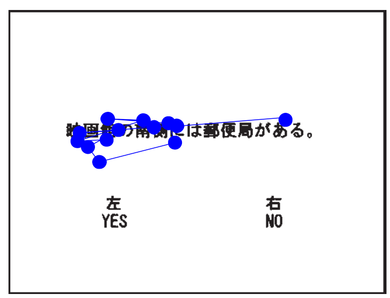


Figure 3. An example of scan-paths

behavior analysis of viewers or users may be possible by using their eye movement data. In this paper, a contextual understanding and transition among internal mental states has been employed, and both understanding and the decision making processes are analyzed.

## III. Experiment

### A. Experimental task

First, participants were asked to understand and memorize a number of definition statements which describe locational relationships between two objects (Figure 1a). Each definition statement was presented for 5 seconds. Second, ten questions were given in statement form, to determine the

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-260-8

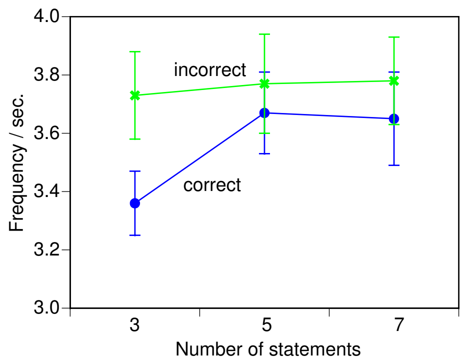


Figure 4. Frequency of saccadic eye movements by response correctness.

degree of understanding. These questions asked participants to select one of two choices about whether each question statement was "Yes (True)" or "No (False)" as quickly as possible (Figure 1b). Each question statement was shown for 10 seconds. All texts were written using Japanese Kanji and Hiragana characters, and the texts were read from left to right. This is a Prolog-like test for human subjects [12]. If the participant can understand the context of the definition statements, they can make the correct decision. No feedback about response accuracy was given. At definition statements 3, 5 and 7 the difficulty of the definition statements increased because the amount of information to be memorized increased. Again, during the experimental session, a set of definition statement containing a number of statements (3, 5, or 7), is presented as shown in Figure 1a, and the 10 question statements in Figure 1b are presented after that. This experimental condition was assigned randomly to participants, to prevent any sequential effects. Five sets of statements were created for each of the three task levels. In total, there were 150 data responses divided into 15 task sets. The subjects were 6 male university students ranging from 23 to 33 years of age. They had normal visual acuity for this experiment.

The correct rates and mean reaction times of responses for question statements are summarized in Figure 2 [3]. As shown in the figure, the response accuracy decreases with the number of definition statements, and mean reaction times depend on whether the responses are correct or incorrect.

### B. Eye-movement measuring

The task was displayed on a 20 inch LCD monitor positioned 60 cm from the subject. During the experiment, participant's eye movements were observed using a video-based eye tracker (EMR-8NL). Eye-movement was tracked on a 640 by 480 pixel screen at 60 Hz, and was recorded on a PC as time course data, while participants read and understood the content of each statement.

The tracked eye movement data was extracted for the period of time participants viewed each question statement

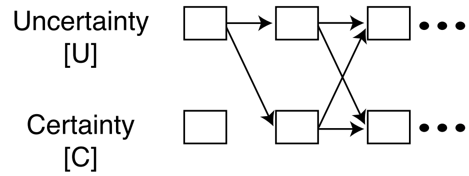


Figure 5. Trellis diagram of a transition.

before they pressed the mouse button. The tracking data was converted into visual angles according to the distance between the participants and the display. Differences between viewing positions were calculated using the recorded time course data, and eye movements were divided into saccades and fixations using a threshold of 40 degrees per second [13].

In Figure 3, eye movements including fixation and saccade are superimposed on a question statement. The dots indicate fixation points, and the lines indicate saccade paths. The appearance of fixations and saccades are related to the correctness of the answers. The frequency of saccades per seconds are summarized in Figure 4. The horizontal axis indicates the number of definition statements, and the vertical axis indicates the saccade frequency. The frequencies are summarized by correct responses and incorrect responses. As shown in the figure, there is a significant difference between answer correctness [14]. Also, selected features of eye movements, such as saccadic eye movements, have been discussed previously [2], [3]. The eye movements are observed at a sampling rate of 60 Hz, and features of eye movement (saccades or fixations) appear in every sample. In this paper, we suppose that the viewer has certainty about an answer when the response is correct. Otherwise, the viewer has uncertainty when the response is incorrect. Though the viewer's certainness about an answer may be fixed at the time of responding, the certainness is not fixed during the thinking about the answer. According to Figure 3, viewers repeatedly read the question statements. Therefore, their certainness about the answer may change before their final response. Here, two states of certainness about answers can be defined as "Certainty" and "Uncertainty", and the level of certainness frequently moves between the two states. As Figure 4 shows that eye movements reflect answer correctness, eye movements may also suggest the state of certainty during the reading of question statements.

## IV. HIDDEN MARKOV MODELING

According to the correct and incorrect responses in the task results, it is supposed that there are two internal states used when making choices. As mentioned in the previous section, these two internal states are defined as "Certainty" and "Uncertainty". The state transition is illustrated as a Trellis diagram in Figure 5. Initially viewers are in a state of "Uncertainty". Then a transition to "Certainty" occurs after

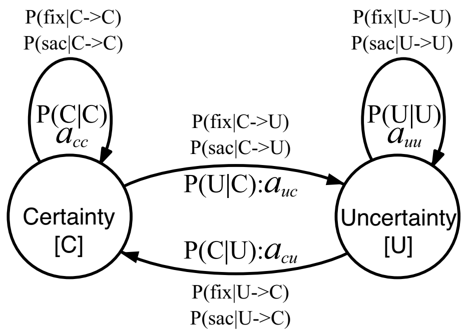


Figure 6. HMM diagram.

that. Also, eye movements may be related to the transition, therefore, specific eye movements such as outputs happen when these state transitions occur.

The outputs are simply saccades (sac) and fixations (fix), as mentioned in the above section. The series of observed eye movements, as the set of time series data, $O$ can be defined using the following formula:

$$\begin{aligned} O &= \{o_1, o_2, o_3, \cdots, o_t\} \\ o_t &= \text{"}sac\text{"} : saccade \vee \text{"}fix\text{"} : fixation \\ t &= T \times 60 - 1, T : sampling\ time(sec.) \\ & \quad 60 : sampling frequency(Hz) \end{aligned} \tag{1}$$

In this paper, a hidden Markov model (HMM), which has two states of certainty for contextually understanding question statements, is employed as a dynamic arithmetic model [7]. A diagram of the model is illustrated in Figure 6. The two states indicate levels of high or low certainty for two responses know as Certainty and Uncertainty. When the internal state changes during eye movement, the probability of a correct response is high when the internal state stays high ("Certainty"), and the probability of such a response is low when the state stays low ("Uncertainty"). When the question statements are displayed, the certainty must be low, thus the first transition is movement from a low to high state of certainty. The transition is illustrated as a trellis diagram in Figure 5. The transition happens for every sample, as shown in the figure.

As a result, the model $\lambda$ can be defined as follows:

$$\begin{aligned} \lambda &= \{S, Y, A, B, \pi\} \\ S &= \{s_i | \text{"}C\text{"} : certain \vee \text{"}U\text{"} : uncertain\} \\ Y &= \{fix, sac\} \\ A &= \{a_{ij}, i, j = C, U | a_{cc}, a_{cu}, a_{uu}, a_{uc}\}, \sum_j a_{ij} = 1 \\ B &= \{b_{ij}(k), i, j = C, U, k = fix, sac\}, \sum_k b_{ij}(k) = 1 \\ \pi &= \{\pi_C = 0, \pi_U = 1\} \end{aligned} \tag{2}$$

The HMM $\lambda$ can be measured using a series of output symbols $O$, as mentioned above. A set of parameters, $\theta \equiv (A, B)$ is optimized using experimental data. The Baum-Welch algorithm, which uses a likelihood function, is employed as shown in (3) [7].

$$\log P(S|O, \theta) \tag{3}$$

The Forward algorithm provides a series of state transitions $S$, which maximize the likelihood function [7].

The performance of predicting the states and responses was evaluated using the leave-one-out technique. One set of responses, such as a data set for one level of difficulty for one subject, was used as a test set, while the remainder of the data was assigned as training data, to optimize the set of parameters in $\theta$. Using the model and a series of output symbols of eye movements, state transitions were simulated. As previously stated, correct and incorrect responses were predicted, and the estimation accuracy was also evaluated during the time series.

In Figure 6, circles represent two states, which are levels of high and low certainty. The probability ($Pr$) of remaining in one state can be defined as $c$ for "certain" and $u$ for "uncertain", while $Pr(c) + Pr(u) = 1$. According to the features of the Markov transition, the transitional probability can be calculated as shown in (4). Then, the probability of certainty can be stated as a time series.

$$\lim_{m \to \infty} \begin{bmatrix} a_{cc} & a_{uc} \\ a_{cu} & a_{uu} \end{bmatrix}^m \times \begin{bmatrix} c^{ini.} \\ u^{ini.} \end{bmatrix} \to \begin{bmatrix} c \\ u \end{bmatrix} \tag{4}$$

## V. Results

The results of the simulation were summarized as a contingency table in Table I, for three of the definition statements, at 550 msec. (0.55 sec.) after onset of the question statement. Here, when a transition remains in a state of certainty at that time, the response may be correct. When the state of uncertainty occurs at that time, the response may be incorrect. According to this hypothesis, the accurate predictions are certainty-correct and uncertainty-incorrect. The prediction accuracy is given as 83.3% ((242+8)/300). The accuracy of the time series is calculated using the same procedure. The results for each condition between 0 and 1 second of the commencement of the statement presentation are summarized in Figure 7. In the figure, "3 tasks" means the prediction accuracy when three definition statements are used.

The accuracy when three definition statements are used is almost always the highest. The accuracy decreases as the number of definition statements increases from 3 to 7. The accuracy of the responses coincides with the level of

 ISBN: 978-1-61208-260-8 

Table I
PREDICTION ACCURACY OF 3 STATEMENTS AT 550 MSEC.(ACCURACY=83.3%)

| | Prediction | |
|---|---|---|
| | Certainty | Uncertainty |
| Correct | 242 | 41 |
| Incorrect | 9 | 8 |

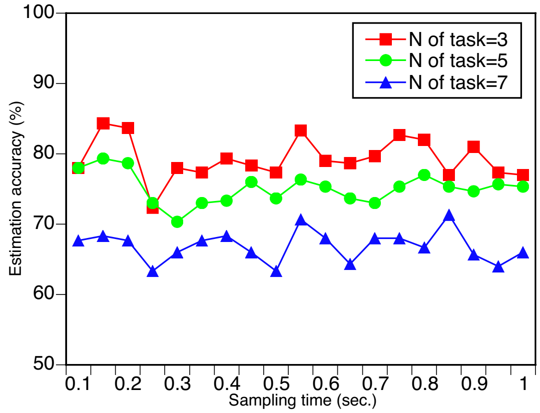


Figure 7. Temporal changes in prediction accuracy.

prediction accuracy. The procedure can not be concerned with the rate of correct responses because the training data consists of a series of equation symbols representing eye movements. According to the changes in prediction accuracy during observation, the accuracy for three statements is high from 0-250 msec. Also, the accuracy for all conditions is high around 550 msec. The highest accuracies, observed at 550 msec., are 83.3% for three statements, 76.3% for five statements, and 70.7% for seven statements. This 550 msec. produces a local accuracy maximum.

The accuracy performance is equivalent to the rates of estimation using all eye-movement data from before the moment when participants responded [3]. Time series prediction produces many response accuracy prediction benefits. During the reading process, accuracy depends on the duration of the decision making process. When eye movements are observed in real time, the response accuracy can be estimated.

According to (4), the probability of remaining in a state of certainty ($Pr(c)$, at $m = t$) can be calculated using the time sequence as well as the sampling rate ($t$ in (1)). Again, the possibility of a correct response is also high when the probability is high at the time. The average probabilities for three levels of task difficulty in the time course are illustrated as bar graphs in Figure 8. The times ($t$) are between 0 and 1 seconds, and also at 2, 3, and 4 seconds. As shown in Table I, the distributions of both responses and predictions shift to the certain state. Therefore, all probabilities are higher than 0.7, but the changes in probability indicate the sensitivity of correct responses. The probability is high from 100-250

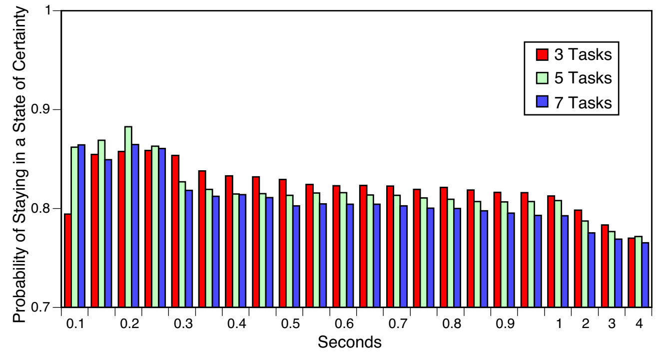


Figure 8. Change in probability of staying a state of certainty by durations.

msec., and decreases with the duration. In particular, the rates decrease gradually after 1 second. The mean reaction times during the experiment are around 3 seconds, and the results indicate that the probability of certainty decreases along with the process of reading the statements. These results may suggest that understanding and decision making are occurring at an early stage, prior to decision responses.

## VI. DISCUSSION

Both prediction accuracy and the probability of remaining in a state of Certainty are affected by the high rate of correct responses. The distribution of training data sometimes shifts to the Certainty state. However, the results for the condition of 7 tasks showed a similar tendency, though the correct rate was smaller than 50%. Since the deviations of the estimation accuracy and the probability along the time course are observed, the model can simulate the transitions of internal states.

The temporal changes in prediction accuracy and the probability of certainty may be concerned with the process of reading statements, because this task requires participants to understand question statements and to recall knowledge memorized from the definition statements. The reading process is tracked using eye movements [6], but the analysis is performed step by step. A typical analysis of the process of reading statements uses event related potentials (ERP). The negative peak is at around 400 msec. (N400) after statement presentation has been affected by the context [15]. This result suggests that statement reading requires more than 400 msec. The local maximum peak of prediction accuracy, concentrated around 550 msec., is included in the same period as the N400 peak. The results of both character perception and eye movement during reading suggest that 150-200 msec. is required to understand the first parts of the statements [6], [16].

According to the evidence, it is possible to suppose that contextual understanding is made between 150 and 400 msec. and that correct or incorrect responses are made after

 ISBN: 978-1-61208-260-8 

that. When the response is delayed, the level of certainty decreases with the duration, because the viewer can not make a certain decision or hesitates to respond. Further examination of these processes will be a subject of our further study.

In this paper, a recall test was employed to measure transitions from an internal state of "Certainty" to "Uncertainty" and back, using eye movements. This basic approach can be applied to general user interface issues. If the internal states are defined as sets of "Good usability" and "Poor usability", then "Like" and other selectable icons, eye movements may indicate a change in user's intentions while the user is interfacing with an application. The examination of this possibility will also be a subject of our further study.

## VII. Conclusion

The hidden Markov model, which consists of two hidden states, has been created to predict the correctness of answer choices in response to questions in recall tests. The model was optimized using observations of participant's responses and their eye movements. The prediction accuracy was calculated using the sequence of eye movement data, and the accuracy decreased with the difficulty of the task. The highest accuracy was observed at 550 msec. after stimuli onset. The probability of choosing the correct response while remaining in a state of certainty was calculated using the probability of transition. The certainty was the highest from 100-250 msec. after stimuli onset, and decreased as the duration increased. The results of the simulation coincided with the experimental observations. These results provide evidences that the model presents a possibility of conducting a dynamic or time series analysis.

## Acknowledgement

This research was partially supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Challenging Exploratory Research (23650528: 2011-2012).

## References

[1] Y. Zhang, *An Item Response Model on Probability-Testing (in Japanese)*. Tokyo, Japan: University of Tokyo Press, 2007.

[2] M. Nakayama and Y. Takahasi, "Estimation of certainty for responses to multiple-choice questionnaires using eye movements," *ACM TOMCCAP*, vol. 5, no. 2, p. Article 14, 2008.

[3] M. Nakayama and Y. Hayashi, "Estimation of viewer's response for contextual understanding of tasks using features of eye-movements," in *Proceedings of ACM Symposium on Eye-Tracking Research & Applications (ETRA2010)*, A. Hyrskykari and Q. Ji, Eds. New York, USA: ACM, 2010, pp. 53–56.

[4] J. H. Goldberg and J. C. Schryver, "Eye-gaze determination of user intent at the computer interface," in *Eye Movement Research, Volume 6: Mechanisms, Processes and Applications (Studies in Visual Information Processing)*, J. Findlay, R. Walker, and R. Kentridge, Eds. Elsevier, 1995, pp. 491–502.

[5] R. Bednarik, H. Vrzakova, and M. Hradis, "What do you want to do next: A novel approach for intent prediction in gaze-based interaction," in *Proceedings of ETRA 2012: ACM Symposium on Eye-Tracking Research & Appliations*, J. B. Mulligan and P. Qvarfordt, Eds. New York, USA: ACM, March 2012, pp. 83–90.

[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] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, USA: Springer Science+Business Media, 2006.

[8] A. Liu, "What the driver's eye tells the car's brain," in *Eye Guidance in Reading and Scene Perception*, G. Underwood, Ed. Elsevier, 1998, pp. 431–452.

[9] Y. Puolamäki, J. Salojärvi, E. Savia, J. Simola, and S. Kaski, "Combining eye movements and collaborative filtering for proactive information retrieval," in *Proceedings of ACM-SIGIR 2005*, A. Heikkil, A. Pietik, and O. Silven, Eds., ACM. New York, USA: ACM Press, 2005, pp. 145–153.

[10] R. J. K. Jacob and K. S. Karn, "Eye tracking in human–computer interaction and usability research: Ready to deliver the promises," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, Hyona, Radach, and Deubel, Eds. Oxford, UK: Elsevier Science BV, 2003.

[11] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *Industrial Ergonomics*, vol. 24, pp. 631–645, 1999.

[12] W. Takita and M. Nakayama, "Influence of audio-visual presentation style for sentences on their understanding," *IEICE Technical Report*, vol. 104, no. 132, pp. 17–22, 2004.

[13] Y. Ebisawa and M. Sugiura, "Influences of target and fixation point conditions on characteristics of visually guided voluntary saccade," *The Journal of the Institute of Image Information and Television Engineers*, vol. 52, no. 11, pp. 1730–1737, 1998.

[14] M. Nakayama and Y. Hayashi, "Feasibility study for the use of eye-movements in estimation of answer correctness," in *Proceedings of COGAIN2009*, A. Villanueva, J. P. Hansen, and B. K. Ersboell, Eds., 2009, pp. 71–75.

[15] M. Kutas and S. A. Hillyard, "Reading senseless sentences: Brain potentials reflect semantic incongruity," *Science*, vol. 207, pp. 203–208, 1980.

[16] H. Abe and M. Nakayama, "Event-related potential study of Kanji perception process," *The Journal of the Institute of Image Information and Television Engineers*, vol. 60, pp. 397–404, 2006.


ISBN: 978-1-61208-260-8