Estimation of Certainty for Responses to Multiple-Choice Questionnaires Using Eye Movements

MINORU NAKAYAMA and YOSIYUKI TAKAHASI
Tokyo Institute of Technology

To examine the feasibility of estimating the degree of strength of belief (SOB) of responses using eye movements, the scan paths of eye movements were analyzed while subjects reviewed their own responses to multiple choice tasks. All fixation points of eye movements were classified into visual areas, or cells, which corresponded with the positions of answers. Two estimation procedures are proposed using eye-movement data. The first one is identifying SOB using scan-path transitions. By comparing subject's reports of high and low SOB and eye-movement estimations, a significant correct rate of discrimination of SOB was observed. When the threshold of discrimination was controlled, a high rate of correct responses was obtained if it was set at a low level.

The second procedure is conducting SOB discrimination using support vector machines (SVM) trained with features of fixations. Subject's gazing features were analyzed while they reviewed their own responses. A discrimination model for SOB was trained with several combinations of features to see whether performance of a significant level could be obtained. As a result, a trained model with 3 features (which consist of interval time, vertical difference, and length between fixations) can provide significant discrimination performance for SOB.

These results provide evidence that strength of belief can be estimated using eye movements.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; H.5.2 [**Information Interfaces and Presentations**]: User Interfaces—*Evaluation/methodology*

General Terms: Human factors

Additional Key Words and Phrases: Eye-movements, certainty, scan-path analysis, support vector machines

ACM Reference Format:

Nakayama, M. and Takakhashi, Y. 2008. Estimation of certainty for responses to multiple-choice questionnaires using eye movements. ACM Trans. Multimedia Comput. Comm. Appl. 5, 2, Article 14 (November 2008), 18 pages. DOI = 10.1145/1413862.1413867 http://doi.acm.org/10.1145/1413862.1413867

1. INTRODUCTION

Common belief is a key concept for communication between humans and with machines and devices; we feel certainty or uncertainty when the degree of shared common belief is strong or weak during communication [Iwahashi 2003b; 2003a; Ishizaki and Den 2001]. During a dialogue consisting of human conversation or human-computer interaction, a person feels certainty in the sharing of a common belief if there is good communication. Otherwise a person feels uncertainty. This suggests that the performance

Authors' addresses: CRADLE (Center for Research and Organization of Educational Technology, Tokoy Institute of Technology, Ookayama, Meguru-ku, Tokyo 152-8552, Japan.

of memory and learning are influenced by certainty and belief, as they are by communication [Imai and Takano 1995].

Certainty may help us to understand human behavior, such as when we search for information while browsing the Internet. Some Internet web sites that provide users with information ask them to provide a response that evaluates the usefulness of the site [Google 2007]. This system is very similar to the relevance feedback of information retrieval systems [Tokunaga 1999]. Certainty pertains to behavior in relation to the difficulty of a task, such as the operation of a computer. When the task is difficult and the certainty is low, we hesitate to operate the computer. Otherwise, we can operate it when the certainty is high. Evaluations, such as system usability tests, are mostly conducted with questionnaire surveys [Brooke 1996]. Recently, the use of online testing is growing, due to the development of online learning. Most online tests estimate the ability of a learner by using Item Response Theory (IRT) [Zhang 2007]. A survey of self-confidence of answers can be used to help develop an IRT model which considers the problem solving response process [Shigemasu and Ueno 1993]. Because self-confidence of answers correlates with correctness [Shinotsuka 1993], certainty of responses may help to create a precision response model. While measuring certainty is useful for various applications, further surveys or interviews are required, however. Additionally, the cost to obtain certainty is not low.

There is the possibility of estimating certainty from human behavioral information. We often try to voluntarily find further information when there is poor communication or we feel uncertainty, and a typical human reaction is the unstable eye movements associated with uncertainty [Underwood 2005]. This shows the possibility of making subjective evaluations of viewers using eye movements. Most eye-movement analyses are based on scan-path analysis [Noton and Stark 1971], therefore irregular eye movements may present typical scan paths. The scan paths based on the fixation points can be noted as strings of fixation region labels. This method allows quantitative comparison of the similarity of scan paths [Choi et al. 1995]. In these approaches, all scan paths and transitions are analyzed to extract low "strength of belief (SOB)," because viewer's uncertainty affects the scan-path pattern. However, the feasibility of measuring certainty using eye-movement patterns is not clear.

The phenomenon has already been applied to detecting relevant text during an information retrieval experiment [Puolämaki et al. 2005]. Poulamäki et al. have applied the Hidden Markov Model (HMM) to detecting relevant text using eye-movement data during document retrieval experiments. They extracted some features of eye movements and analyzed them. These extractions were based on some of the characteristics of eye movements while viewers were looking at a text [Salojärvi et al. 2004]. As the HMM is based on the probabilistic transition function within a finite set of states [Stork et al. 2001], it is not easy to create a discrimination model using appropriate transition states [Tsuda 2003], however. The possibility that discrimination performance using support vector machines (SVM) is comparable to the performance using HMM has also been suggested [Shimodaira et al. 2001]. The building of a discrimination model using SVM to obtain appropriate performance is easier than using HMM when the training data, consisting of examples of features and desired target categories, is provided [Tsuda 2003].

In this article we examine the feasibility of estimating an index of the SOB for answers to multiple choice tasks [Nakayama and Takahasi 2006] by using eye movements for fixation points, where subjects were looking at answers they had chosen in a previous session of the experiment. The detailed aims are as follows:

(1) To examine whether eye-movement based certainty can be an index of the (SOB) for answers to a multiple choice task.

(2) To develop a procedure for determining SOB for a fixation area using scan-path analysis and to evaluate that performance.

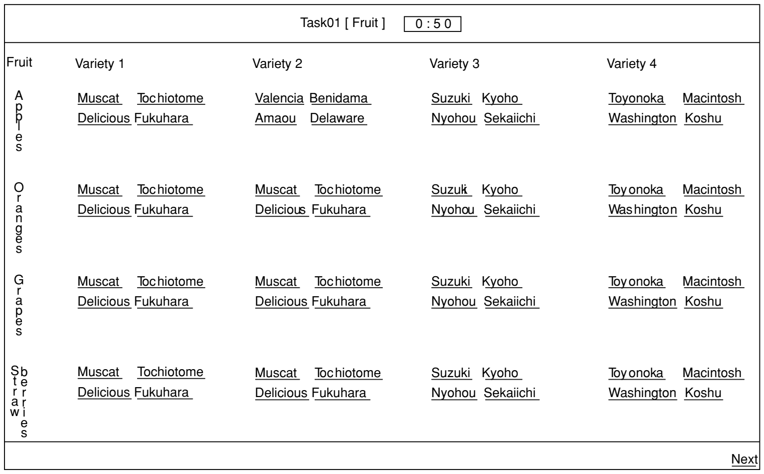

Fig. 1. Question display for the answering session (English translation).

(3) To develop a method of estimating SOB for a fixation area using Support Vector Machines with features of eye movements and to evaluate that performance.

Toward these aims, we have developed experimental tasks and analyzed eye movements, after which we proposed two methods to estimate viewer's SOBs.

2. EXPERIMENTAL METHOD

2.1 Experimental Task

A set of 16 multiple choice tasks, arrayed in a 4 by 4 matrix of 4 questions each with 4 answer choices, was prepared as a full-screen Web page. A screen shot of the question display for the answering session is shown in Figure 1. While this figure shows an English translation, the experiment was conducted using a Japanese version. There is no difference between the presentation formats in English and Japanese. The texts are read from left to right in both cases, though Japanese text is usually written vertically. The category of the fruit on the left-hand side was displayed as vertical text in the Japanese version. This multiple-choice task required the subject to select an appropriate response from the 4 answer choices using a computer mouse. The multiple choice questions were arranged 1 question to a row, 4 rows to a column, as shown in Figure 1. The question appeared in the leftmost column, and four sets of 4 candidate answers appeared in the four columns on the right-hand side. Each question gave a type of fruit (apples, oranges, grapes, and strawberries in this example), and required the subject identify which answer of the four was a variety of that type of fruit. The names of some varieties are local Japanese names, and are easily recognized by subjects. Most are popular and displayed at fruit shops in Japan, therefore subjects can easily categorize them. Some minor varieties were added to control total task difficulty, however. For the question about Apples in the first row, the candidate answers were Muscat, Tochiotome, Delicious, and Fukuhara, four varieties of grapes, strawberries, oranges and apples respectively. Looking at the column titled Variety 1, subjects had to identify which was a variety of that type of fruit by clicking on, in this case, Fukuhara, a variety of apple. When a subject selected one of the four choices, the three choices not selected disappeared. A screen shot of the display for the reviewing session is shown in Figure 2. If a subject wanted to change an answer during the answering session, the initial screen could be returned to by clicking "return."

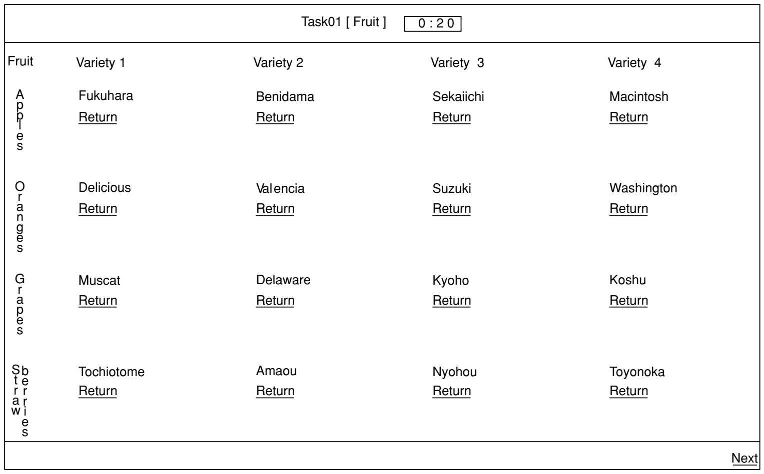

Fig. 2. Answer chosen display for the reviewing session (English translation).

The content of the questions was selected in accordance with the results of a preliminary experiment. The task consisted of an answering session and then a reviewing session.

2.2 Experimental Procedure

The experiment was conducted using consecutive answering and reviewing sessions. One minute was assigned for both sessions. The procedure is shown as follows:

(1) Practice session for answering sample questions
(2) Calibration for eye tracking
(3) Answering session (Figure 1)
(4) Reviewing session (Figure 2)
(5) Printout of screen shot
(6) Declaration of a rate of strength of belief (SOB) for each question
(7) Return to step 2

An experimental setup is shown in Figure 3. The setup is a typical browsing session using a PC without any measuring of eye movements. Measurement details will be described in later sections. Five subjects participated in both sessions. For the first minute of the experiment, answers were selected, followed in the second minute by the reviewing session, where subjects reviewed their own answers without making corrections. The times for answering and reviewing were strictly controlled by a computer timer. A practice session was prepared so that all participants throughly understood the experimental procedure.

In this experiment, we examined the relationship between SOBs and eye-movement behavior in the reviewing session. Each subject reviewed three sets of tasks so that in total 48 SOBs were reported.

2.3 Rating Strength of Belief (SOB)

Subjects freely noted their own subjective certainty for each answer of 48 in total (16 tasks $\times$ 3 sets) as self-confidence, or SOB, on a scale between 0 and 100 using a pen and a copy of a screen shot of the selected answers. If the subject had strong self-confidence in an answer, SOB marks approached 100, otherwise the SOB mark was near 0 when the subject did not have any confidence. Score distributions

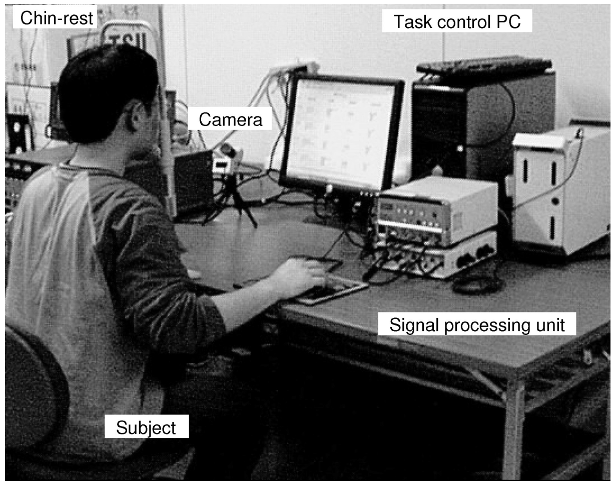

Fig. 3. An experimental setup.

of SOBs depended on subjects, therefore. This subjective evaluation is usually conducted after the reviewing session, according to the provisions of the experiment in Section 2.2.

3. RESULTS OF ANSWERING PERFORMANCE

3.1 Relationship Between SOB and Answer Correctness

To examine the response accuracy, the correct rate of answers are calculated for each subject. The total correct rate is 72.0%, and the correct rate for individuals is 62.5 ∼ 76.3%. The standard deviation is 5.7, which is relatively small. This suggests that the task difficulty can be independently controlled amongst subjects.

The SOB for each question was freely reported by subjects as an analog value (0 ∼ 100). To examine the relationship between subjective reports of certainty and answer correctness for questions, the average reported SOB across subjects was summarized as a hit (N = 5, mean = 75.0%, SD = 7.1) or a miss (N = 5, mean = 38.4%, SD = 8.7) in Figure 4. There is a significant difference in SOB between hit and miss responses ($t(8) = 7.3$, $p < 0.001$). This suggests that subject's reports show the correctness of their answers, and this confirms the relationship between self-confidence and the rate of correct responses in a previous piece of research [Shinotsuka 1993].

3.2 Two Classes of SOBs

Since answer correctness can be classified as a hit or a miss for this question type, SOB values (z) of subject's reports were also divided into two levels ($z = \{high\ SOB, low\ SOB\}$) using a threshold to obtain the level of answer correctness. Here, it is supposed that SOBs reported tend to the normal distribution. According to the hypothesis, two distributions of SOB for hit and miss responses can be illustrated as in Figure 5. This figure shows two distributions which are estimated from mean and standard deviations (SD) of SOBs for hit and miss responses. To classify SOBs into two classes for maximizing the two probabilities of *high SOB* for hits and *low SOB* for misses, the threshold can be set as the overlap point of two normal distributions. When the threshold is increased, the rate of hit responses decreases in the *high SOB* category, while the rate in the *low SOB* category increases. To obtain information about the responses, the threshold can be controlled.

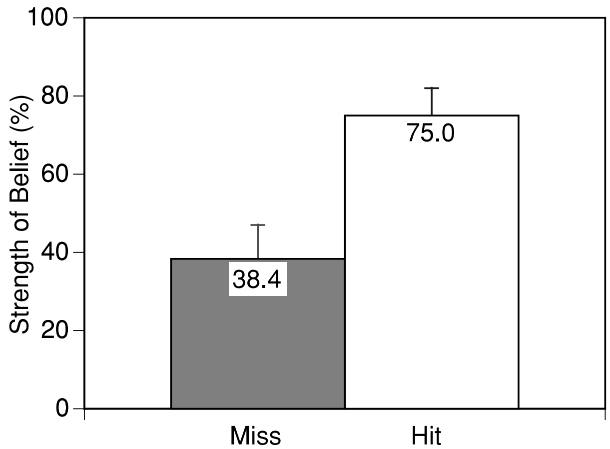

Fig. 4. Mean strength of belief for hit and miss responses.

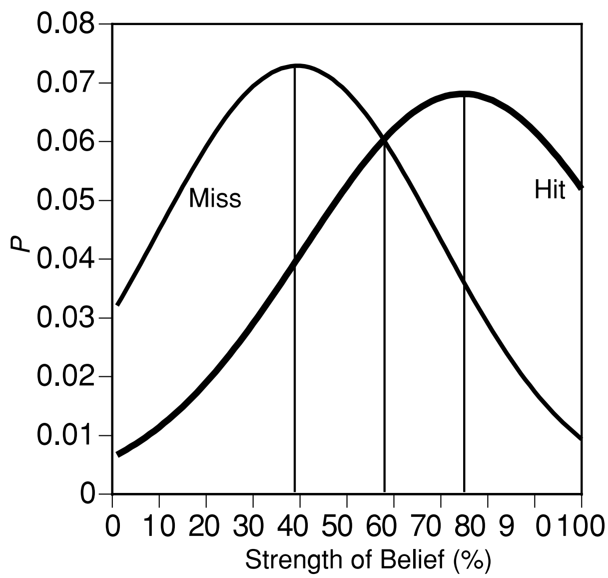

Fig. 5. Distributions of strength of belief for hit and miss responses.

According to this procedure, SOBs are classified by each subject into high and low SOBs using a mean threshold (58.0%) as the overlap point, without considering whether the answer has been a hit or a miss. As a result, 64.6% of all responses are classified as high SOBs, and the rest are classified as low SOBs. The whole SOB distribution is shifted to a high value, as in Figure 5, where the number of reports of high SOBs is more than the number of reports of low SOBs.

4. SOB ESTIMATION BASED ON SCAN-PATH ANALYSIS

4.1 Measuring Eye Movements

During the experiment, eye movements for each subject were observed using a video-based eye tracker (nac:EMR-8NL), as shown in Figure 3. The task was displayed on a 17-inch LCD monitor positioned

Table I. Statistics of Eye Movements Across Subjects

Subject	Fixation Points(*)	Fixation Duration Mean (msec.)	Fixation Duration SD	Saccade Length Mean (deg.)	Saccade Length SD
A	50.7	137.4	42.8	8.6	6.4
B	78.7	179.3	95.8	6.6	4.8
C	62.6	148.2	52.8	7.2	4.9
D	52.6	133.4	42.8	8.7	5.3
E	77.3	162.7	63.9	7.3	5.6
Overall	64.4	155.2	68.0	7.5	5.4

(*) mean points per reviewing session in one minute.
Fixation was extracted using a commercial software.

65 cm from the subject. The subject rested his or her head on a chin rest and a small infra-red camera was positioned between the subject and the monitor, 40 cm from the subject. The subject's hands were always free, so that he or she was not restricted during the task. The tracker was calibrated at the beginning of every session. Eye movements were tracked on a 800 by 600 pixel screen at 60 Hz, and every tracked position was specified in pixel scaled data. Therefore, as a theoretical assumption, spatial resolution is about 0.03 degrees of visual angle per pixel. The accuracy of the spatial resolution of this equipment is noted as 0.1 degrees in visual angles in the catalogue [nac Image Technology 2007], however. Eye-movement data were recorded on a PC as time course data. The tracking data were converted into visual angles according to the distance between the viewer and the display, so that the visual angles of the display were ± 10.5 by ± 13.5 degrees.

Eye movements were divided into saccades and fixations using a commercial analyzing program [nac Image Technology 2004]. A fixation is defined in the software as eye movements staying within a 0.3 degrees visual angle, at a velocity of 3 degrees/sec or less, and 100 msec. as the minimum duration. The detailed parameters were set according to the characteristics of eye movements [JSVS 2001]. It is not easy to measure accurate saccade speeds, however, because the sampling rate for eye tracking of this equipment is largely irrelevant for ballistic eye movements. Therefore, mainly fixation points and fixation durations are analyzed in this article. Basic statistics of eye movements such as mean number of fixation points, mean fixation duration, and mean saccade length during a reviewing session are summarized across subjects in Table I. According to the results, there were 64.4 fixation points per one minute in the chosen answer reviewing session, and the mean fixation duration was 155.2 msec. Although mean fixation duration and SD for Subject B were the longest, the mean fixation difference was not large. According to a previous statistical summary [Rayner 1998], the mean and distribution of duration were smaller than the ones for silent reading (255 msec) and visual search (275 msec), and the mean saccade length (7.5 deg) was longer than the one for scene perception (4 deg). The measuring accuracy and extraction procedure of fixation using this eye-tracker, which are mentioned above, may affect the differences in measurements. Additionally, the reviewing time was only one minute and the chosen answers were distributed across columns and rows, therefore fixation duration was suppressed. Also, saccade length for a chosen answer was longer than the width of the column and the height of the row of any cell. This suggested that in comparing basic characteristics of results from this experiment and a previous study, the differences in measuring conditions should be considered carefully. An examination of Table I reveals several tendencies of the measured data in association with the results of eye-movement behavior, which merit discussion.

During observation, it is assumed that the size of the viewing field is almost constant, so the viewer obtains visual information mainly from the fixation point, which is measured by an eye tracker. To specify the fixation points for question items or chosen answers, all fixation points were classified into 4 cells for question items and into 16 cells for chosen responses to the questions, as shown in the grid

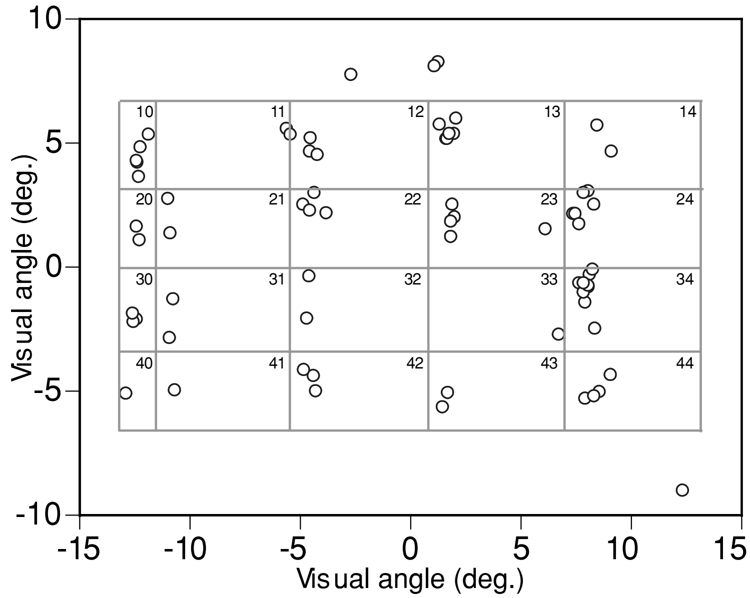

Fig. 6. Fixation point classification into cells of question items and chosen answers.

in Figure 6. The cells were set up in accordance with the positions of the question items and the chosen answers. The fixation cell numbers are marked for each cell in the top-right corner (for question items: 10, 20, 30, 40; for cells of chosen responses: 11 to 44) according to the definition mentioned later in the article. The rest areas were headers and footers of the display, including the OS message bar area on the screen. Therefore, scan paths can be noted as a series of fixation cell numbers. In this article, transitions of fixation points across the cells are analyzed.

4.2 Scan-Path Analysis

A sample of scan paths during the reviewing session illustrated as straight lines between fixation points is shown in Figure 7. The scan paths between fixation points were extracted by analyzing the time course data of eye movements. To note the sequence of fixation points across the cells mathematically, the cell q for row I and column J is defined as follows:

$$(q_{ij}, I : i = 1, \ldots, 4, J : j = 0, \ldots, 4, j = 0 : \textit{question item})$$

According to the definition, numerical labels were given to each cell, so scan paths can be presented as strings of labels [Choi et al. 1995]. The region labels were based on different information cells, question items, and chosen answers. In other words, the reviewing process is illustrated in a state transition diagram, where fixation cells are defined as states. To determine whether any eye movements are due to uncertainty, two scan-path series between fixation points, namely the state transitions, were analyzed. As there are many transition patterns across cells, we focused on a transition pattern that is illustrated in Figure 8.

For example, a transition is observed from cell q_{ak} for row $a(a \in I)$ and column $k(k \in J, 1 \leq k \leq 4)$ to same cell q_{ak} or to another cell $q_{bk}, (a \neq b)$ in the same column via question item cell q_{i0}. When a three-cell transition (q_{ak}, q_{i0}, q_{bk}) happened, this transition suggested that a viewer had confirmed his or her own answer or the consistency of his or her answer choices because the certainties for these

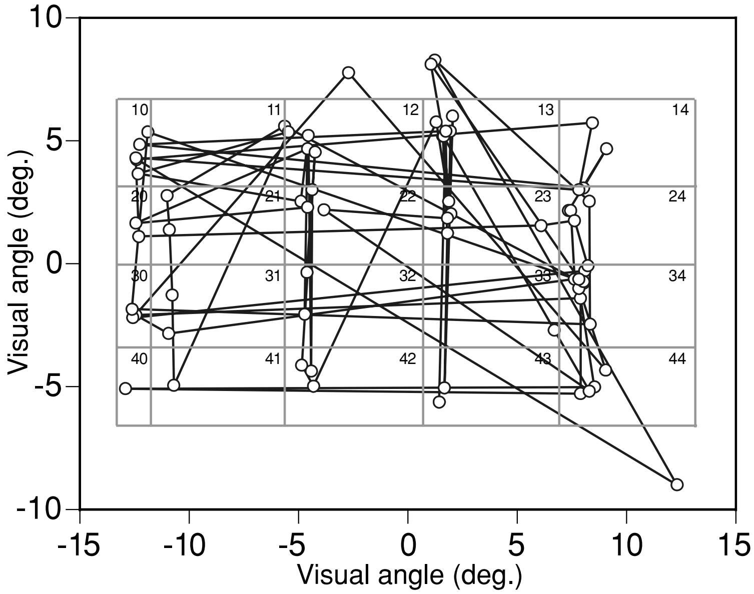

Fig. 7. Scan paths were superimposed on the cells of question items and chosen answers.

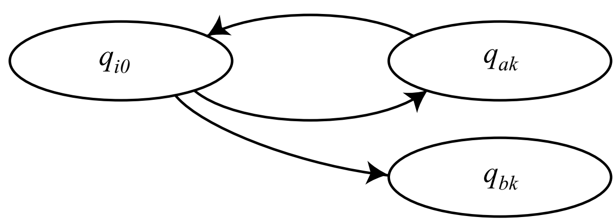

Fig. 8. Transition state diagram for fixation cells.

might be low. This estimation is based on the hypothesis that specific eye movements are associated with uncertainty.

Here, estimated SOB from eye movements is noted as $\hat{z}_{ij}$ for cell q_{ij}, therefore, $\hat{z}_{ij} = \{$*high SOB*, *low SOB*$\}$. All SOBs for the answers cells were set to *high SOB* in the initial condition. SOBs for q_{ak} and q_{bk} were changed to *low SOB* when three fixation transitions occurred. If a viewer gazed at question item cell q_{i0}, the SOB for q_{ak} or both the SOBs for q_{ak} and q_{bk} were recognized as low SOBs. When all other patterns occurred, the SOBs for engaged cells of answers were kept high.

The three typical scan paths illustrated in Figure 9 were extracted from Figure 6. The arrows indicate a transition of fixation points. The first scan path [A] was noted as a three cell transition (q_{12}, q_{10}, q_{22}). Therefore, the scan path fits the transition pattern in Figure 8. Both SOBs for q_{12} and q_{22} change to *low SOB* ($\hat{z}_{12} =$ *low SOB*, $\hat{z}_{22} =$ *low SOB*). The second scan path [B], noted as (q_{31}, q_{30}, q_{34}), does not fit the three cell transition pattern in Figure 8 because the fixation point moved from the first column q_{31} to the fourth column q_{34} via question cell q_{30}, and both SOBs for q_{31} and q_{34} were kept high ($\hat{z}_{31} =$ *high SOB*, $\hat{z}_{34} =$ *high SOB*). The third scan path [C] showed a recursive cell transition: (q_{44}, q_{40}, q_{44}). SOB for q_{44} changed to low ($\hat{z}_{44} =$ *low SOB*).

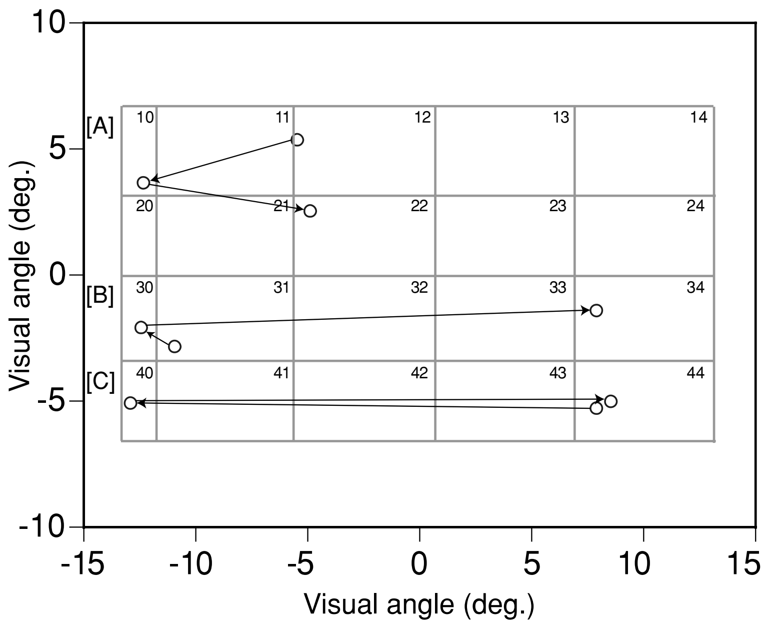

Fig. 9. Three typical scan paths in a reviewing session.

Table II. Contingency Table for Estimations

Eye-movement estimation ($\hat{z}$)	Subject's report (z) SOB [High]	SOB[Low]
SOB[High]	Correct	Miss
SOB[Low]	Miss	Correct

These two-class classifications were automatically calculated from the scan-path data by a computer program using a three-cell transition model. This is an estimation method for SOB using scan-path patterns.

4.3 Estimation Results based on Scan-Path Analysis

The eye-movement patterns were analyzed using the model of fixation cell transition shown in Figure 8. In the outcome, the average number of transitions across the cells on the screen being reviewed was 61.1 within defined cells, and the average number of fixation cells for question item q_{i0} was 4.8. By applying the three cell transition model in Figure 8, an average of 2.9 out of 16 questions were assigned to *low SOB*.

Subject's SOB reports and eye-movement estimations were summarized in a contingency table (Table II). When the estimation coincided with two classes of SOBs based on a subject's report, it showed that eye movements can estimate SOB correctly. Estimation results using three cell transition models are summarized in Table III (threshold = 58.0%), using the same format as Table II. According to the analysis, the total rate of correct responses was 68.8%, and was significantly higher than chance ($p < 0.05$). The rate of correct responses for *high SOB* is the highest. The estimation performance for *low SOB* is low, because more than half of the low SOBs are mistakenly categorized as *high SOB*. Therefore estimation performance is very weak for *low SOB*, and it may depend on applying an estimation algorithm which uses eye-movement scan paths.

Table III. Results of Discrimination for Strength of Belief using Scan-Path Discrimination (%)

Eye-movement estimation ($\hat{z}$)	Subject's report (z) SOB [High]	SOB[Low]
SOB[High]	57.5	24.1
SOB[Low]	7.1	11.3

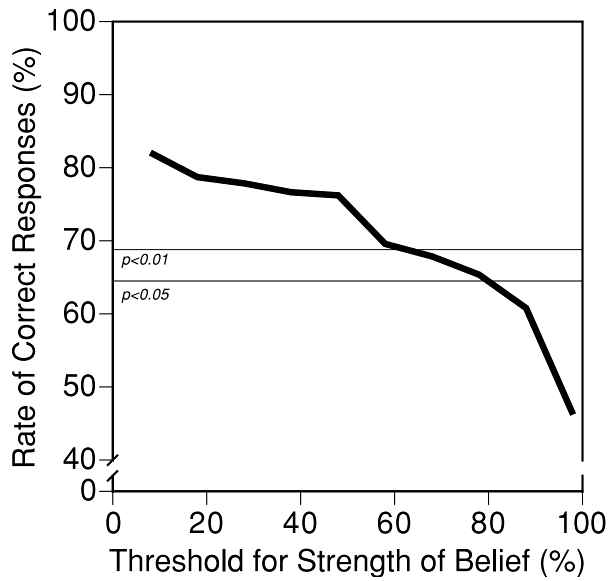

Fig. 10. Rate of correct estimation against the threshold for SOB, using scan-path estimation.

The correct rate of estimation may also depend on the threshold, which is the overlap point between hit and miss responses. Therefore the rate was investigated in accordance with the threshold. The threshold was controlled to change from 58% for the the average threshold across all subjects, in increments of 10% up and down, until 100% and 0% were reached (for example, threshold = 8.0, ..., 48.0, 58.0, 68.0, ..., 98.0 %). When the threshold was adjusted from 0 to 100%, the percentage correct changed with the SOB threshold, as illustrated in Figure 10. Significant levels of estimation of SOB are displayed in the figure.

According to Figure 10, the rate of correct responses decreases monotonically with the threshold for SOB. The estimation performance is better when the threshold is set to a lower value. Figure 10 shows that a significant rate of correct responses is obtained when the threshold is lower than 60%. This result suggests that a scan path between a question item and an answer area appears when an SOB report is low.

This estimation procedure detects only *low SOB* for responses, so that the number of three cell transitions, which are extracted as in Figure 8, is small. Therefore, this estimation procedure may be significant when the threshold is set too low and the focus is *low SOB* response. The performance is low when the threshold is set too high, however. To determine the performance of this estimation procedure, a detailed analysis is conducted as follows. Most SOBs distributed around the mean SOB as shown in Figure 5. The performance was examined with a threshold (58.0%) which was extracted from the overlap point in Figure 5.

Table IV. Results of Recall and Precision Rates (%) Using Scan-Path Discrimination

Eye-movement estimation	Precision	Recall
SOB[High]	70.4	89.0
SOB[Low]	61.4	31.8

4.4 Recall and Precision Rates

Recall and precision rates are often used as evaluation metrics of discrimination performance [Jackson and Moulinier 2002]. A recall rate is defined as the percentage of correct hits, meaning that the estimation of SOB coincides with the subject's report, based on the number of correct reports. The precision rate is defined as the percentage of correct hits per correct estimation. Therefore both rates are derived using the following equations:

$$Precision\ rate = \frac{coincident\ correct\ response}{correct\ response\ based\ on\ eye\ movements}$$

$$Recall\ rate = \frac{coincident\ correct\ response}{correct\ response\ based\ on\ subject\text{'}s\ report}.$$

According to Table II, there are two correct decisions; therefore, recall and precision rates are calculated from Table III as rates for *high SOB* and *low SOB*, respectively. The rates at the threshold are summarized in Table IV. These results show that precision and recall rates for *high SOB* are higher than those for *low SOB*. Both precision rates between high and low SOBs are comparable, but the recall rates between them are significantly different. In particular, the recall rate for *low SOB* is too low, and this shows that the performance for low SOBs is not good.

The algorithm which detects *low SOB* using eye movements may affect the difference in performance between high and low SOB. According to the algorithm, if the initial SOB for all cells is set high, the detection rate of three cell transitions is low. This means that there is an advantage to estimating *high SOB* rather than estimating *low SOB*. Therefore the performance of precision and recall rates for *high SOB* is higher than the performance for *low SOB*.

5. CERTAINTY DISCRIMINATION USING FEATURES OF FIXATION

An estimation procedure for SOBs using scan paths, which is based on the detectability of three-cell transitions, has some restrictions. For example, extraction of three-cell transitions is needed, then an estimation of scan paths for question items is required, and the estimation can only be conducted for *low SOB*. Therefore, estimation is impossible when scan paths for question items have not been observed. To prevent this dependency, the estimation should be conducted using some features of eye movements without scan-path information.

The second procedure is designed to conduct discrimination using temporal and spatial information on fixation points. SOB prediction can be defined as a discrimination problem between two classes of SOBs, based on the features of fixation and eye movements. The support vector machine (SVM) method is a possible discrimination tool for SOBs using features of fixation points, as noted in the introduction.

The discrimination procedure and its performance are examined in the following section.

5.1 Features of Fixation as Eye Movements

Features of eye movements were extracted as follows. We looked at fixation points and fixation patterns as well as scan paths.

All fixation points can be defined as p_k with the time series number k, and they were extracted from the algorithm of eye movements mentioned in Section 4.1. Each fixation point has several attributes as temporal and spatial information: horizontal and vertical positions in the visual angles, and the fixation duration (*fd*). Fixation positions may not be affected by SOB because they depend on the task style and layout. To consider the interval characteristics of fixations as an extension, we looked at the differences between two successive fixation points (p_k, p_{k+1}). These were the horizontal and vertical differences (*dx, dy*), and the length of the two points (*length*) as a saccade length. The results of eye-movement analyses suggest that this eye movements category includes saccadic movements and smooth pursuit movements, because some interval times between the two fixation points are too long to be ballistic movements [Palmer 2000]. Therefore, the interval times (*it*) may show behavior for acquiring further information, and they are affected by the fixation point of the eye when reviewing the preceding answer chosen. All of these characteristics can be considered to be features of a fixation point. They are presented as forming a feature vector for a fixation point.

As a result, feature vectors (V_k) for fixation points (p_k) were produced and noted using the following 5 components [Nakayama and Takahasi 2007]:

$$V_k(fd(fixation\ duration), it(interval\ time), dx, dy, length).$$

5.2 Training Procedure

To discriminate between *high SOB* and *low SOB* with feature vectors using SVM, prototypical training data were produced in the following manner. When a fixation point p_k is included in a cell of chosen answer q_{ij} ($p_k \in q_{ij}$), the feature vector V_k may be reflected by the SOB of the cell z_{ij} ($z = \{high\ SOB, low\ SOB\}$). Here, the SOB value ($t_k$) was given to two classes of SOBs as $t_k = \pm 1 (+1 : high\ SOB, -1 : low\ SOB)$. Therefore, the problem is defined as the prediction of the SOB value t_k from the feature vector V_k of the fixation point in q_{ij}. As a result, the training data were produced as sets of (V_k, t_k).

The problem of data classification was solved by using training data (V_k, t_k) derived with LIBSVM [Chang and Lin 2001] as a tool for SVM training and simulation. In the training procedure, the penalty parameter of the error term C for the soft margin, and the γ parameter as a standard deviation of the Gaussian kernel, were optimized. The data were successfully classified into correct SOB classes using a model trained with the data of all subjects at once. Therefore, the performance evaluation was conducted as a Leave-one-out method. To examine estimation performance for a subject's set of answers, a model was trained using all other data, and estimated SOBs for a set of answers were predicted using the trained model.

After the training, estimated SOB $\hat{t}_k$ for a feature vector V_k was given using a trained SVM model G as $\hat{t}_k = G(V_k)$. The predicted SOB $\hat{z}_{ij}$ for each cell q_{ij} is derived as a logical sum (OR operation) for all $\hat{t}_k$ for $p_k \in q_{ij}$. If there was no fixation point in a cell, the SOB for the cell is defined as *high SOB*. This means that the SOB for a chosen answer is high because the subject does not mind the chosen answer. The feature vector (V_k) was modified as a combination of vector components, and training and testing were conducted.

5.3 Results and Discussion

5.3.1 *SVM Discrimination Performance.* After examining the SVM's discrimination performance using appropriate combinations of feature components of the prepared training data, overall performance for the selected combination of vector components is summarized in Table V. This table compares performance using scan-path analysis. As a result, the SVM performance with 3 features (interval time, vertical difference (*dy*) and length of successive fixation points in reviewing) and the SVM performance with more than 3 features in Table V provide a significant overall rate of correct responses (64.2%), and are comparable with the correct percentages using scan-path analysis. This result suggests that SOB for a chosen answer can be estimated from features of fixations in eye-movement data.

Table V. Discrimination Performance Using the SVM Model Across Feature Vectors

Subject	Scan path	2-dim (it, dx)	2-dim (it, dy)	3-dim (it, dx, dy)	3-dim $(it, dy, length)$	4-dim $(it, dx, dy, length)$	5-dim $(fd, it, dx, dy, length)$
A	**64.6**	47.9	56.3	50.0	58.3	50.0	50.0
B	**66.7**	52.1	**77.1**	47.9	54.2	52.1	47.9
C	58.3	60.4	60.4	**64.6**	**64.6**	**64.6**	62.5
D	**79.2**	52.1	52.1	**68.8**	**68.8**	**68.8**	**70.8**
E	**75.0**	54.2	47.9	**77.1**	**75.0**	**85.4**	**87.5**
Overall	**68.8**	53.3	61.3	61.7	**64.2**	**64.2**	**64.2**

Bold figures show significantly correct percentages ($p < 0.05$).

In comparing the performance amongst individual subjects, the performance of each subject depended on the features of their eye movements. When *low SOB* answers were detected using scan-path analysis, the performance of all subjects was significant, except for one. The total rate of correct responses was 68.8%, and was significantly higher than chance ($p < 0.05$). For subject C, the mean of the horizontal difference (dx) was positive (rightward direction), although the means of one of the other subjects were negative (leftward direction). This suggested that subject C had few leftward scan paths towards an item in the question column, so it was not easy to detect the transition in Figure 8. Where the first method of estimation is based on the explicit scan paths, the performance obviously reflects the difference in features. SVM discrimination is conducted using nonlinear transformation of the features, however.

SVM discrimination performance was determined by referring to the combination of the components in feature vector (V_k). The SVM performance with 3 features (interval time, horizontal difference (dx), and vertical difference (dy)) is not completely significant, but the performance of three subjects is. In particular, the percentage correct of Subject C is significant, although the performance is not significant based on scan-path analysis. This tendency is not significant according to the performance with two features (interval time, horizontal difference (dx)). According to Figure 9, which shows typical scan paths for estimating SOBs, the horizontal difference was not explicitly associated with high or low SOBs. Therefore, the contribution of horizontal difference (dx) may not be great. Of course, it must depend on the task and the display design. This point should be considered in the following discussion. On the other hand, performance surely depends on the combination of features, as those dependencies are noted when analyzing performance.

The performance difference for the component combination between (interval time and dx) and (interval time, dx, dy) suggests the significance of the vertical differences (dy). The performance with 2 features (interval time and dy) is not completely significant, but this performance is significant for one subject, namely B. For subject B, it is the only significant performance across the combinations. The performance with 3 features (interval time, dy, and length) is completely significant, as is the performance using scan-path analysis, and most performances for subjects are significant. The overall performance was the same as the rate with 3 features in Table V. The performance with 4 features (interval time, dx, dy, and length) is also completely significant, but the performance of two subjects was poor because the horizontal difference feature was added. The performance with all 5 features (fixation duration, interval time, dx, dy, and length) is also completely significant, but the performance of three subjects was poor, however. This suggests that some features are not significant for discrimination and therefore performance is negatively influenced. This shows again that the selection of features is key for discrimination. It is interesting that vertical differences (dy) are significant features for discrimination, because there are many vertical eye movements for confirming the chosen answers across rows in a series of columns of fruit varieties, as shown in Figure 7. These eye movements may be

Table VI. Results of Discrimination for SOB Using Features $_{(it,\ dy,\ length)}$(%)

Estimation using features ($\hat{z}$)	Subject's report (z)	
	SOB [High]	SOB[Low]
SOB[High]	58.3	29.6
SOB[Low]	6.3	5.8

Table VII. Results of Recall and Precision Rates for Feature Based Discrimination (%)

Estimation using features	Precision	Recall
SOB[High]	66.3	90.2
SOB[Low]	47.9	16.4

directly associated with the SOBs of the chosen answers, so therefore the vertical difference (dy) is more significant than the horizontal difference (dx).

On the whole, some features of uncertainty surely affect eye-movement behavior, but this representation may be different between individuals. The SVM discrimination performance of subject A is not significant across combinations of features of eye movement, however. The training data included features in two out of three experimental sessions, and the SVM needs to learn which common features show the relationship between certainty and eye movements. If there were few common features, for example, Subject A used different strategies to review chosen answers in three sessions, and the trained SVM could not correctly discriminate the SOBs. This may suggest that SVM discrimination can not achieve perfect performance as well as the first method using scan-path analysis.

5.3.2 *Recall and Precision Rates.* To analyze the discrimination performance of high and low SOBs, discrimination results using 3 features (interval time, dy, and length) are summarized as a contingency table in Table VI, according to the format of Table II. This condition was selected because overall performance is the highest and the number of features is smaller than in the other cases with the same performance. The rate of correct responses for *high SOB* is high and is comparable to the rate using scan paths in Table III. The correct rate for *low SOB* is lower than the rate using scan-path analysis, however. This suggests that the SVM model can discriminate *high SOB* correctly, but it classified certainty into *high SOB* even for *low SOB*. Therefore the rate of correct responses for *low SOB* stayed at a low level. For an algorithm to discriminate SOB using scan-path analysis, classifying certainty into *low SOB* depends on the detection performance of three cell transitions. Therefore, discrimination performance for *low SOB* stays at a low level. As the SVM model for SOB was developed to discriminate two classes directly from the fixation features of eye movements, discrimination performance for *low SOB* cannot be improved, however. As the rate of the number of high SOBs is higher than the rate for the number of low SOBs, the usage of training data for high SOBs may be an advantage.

According to the definitions of recall and precision rates, both rates are calculated from Table VI and the results are summarized in Table VII, in the same format as Table VI. These results show that precision and recall rates for *high SOB* are higher than those for *low SOB*. In comparing these with the results in Table IV, the precision rates for *high SOB* are comparable and the recall rate for *high SOB* gets slightly better, but both rates for *low SOB* are lower than the rates using scan-path analysis. The discrimination performance for *low SOB* is low for the reason mentioned above. Therefore, total performance for estimating *low SOB* is worse than for estimating it using scan paths. The method of creating a data set for SVM training will be a topic of our further study.

6. GENERAL DISCUSSION

The possibility of estimating viewer's certainty or the SOB (Strength of Belief) using eye movements is determined using two approaches; scan-path analysis and SVM discrimination using features of fixations. The hypothesis was that eye movements are associated with the viewer's uncertainty. This phenomenon was studied using estimation procedures that examined scan paths and features of eye movements while subjects reviewed their chosen answers under specific experimental conditions. The experimental conditions were designed to evoke typical eye movements in response to the purpose, which were eye movements influenced by SOB. For example, task contents and construction, task layouts such as locations of question and candidate answer choices, procedures used for answering and reviewing were carefully considered. Additionally, the level of SOBs was divided into two classes in order to predict hit or miss responses. Although the experimental procedure was not simple and there were many restrictions in the experimental conditions, the results provided evidence that the SOBs of the viewer could be estimated using eye movements. Again, viewer's uncertainty affected scan paths and features of eye movements for viewing objects. To extract this information, careful design of the experiment's parameters is key for the accurate detection of task execution and viewer's behavior, for example. Therefore, as the results achieved depend on the experimental conditions, it is not easy to generalize using this evidence. To extend estimation possibility to other conditions, the following topics should be considered carefully: Firstly, a definition of the levels of SOBs. Obviously, the behaviors which are associated with degrees of uncertainty should be analyzed and the relationship between them should be examined. In particular, the threshold setting for SOBs seriously affects the overall performance of the estimation, so the measuring procedures for SOBs should be considered carefully. If accurate classification of SOBs is possible, it may also be possible to estimate multilevel SOBs. These issues are based on techniques beyond using eye movements for measuring viewer's responses, such as establishing a threshold level from SOB distribution in Figure 5. Analysis of subjective responses and rating evaluations should be taken into account in advance. Secondly, task creation and definition, and the task feasibility of detecting classes of SOBs should be examined in detail. The task should be designed so that it is possible to capture the behavior of eye movements in response to the level of SOBs.

An approach to SOB estimation using scan paths was based on the detection of eye movements evoked by uncertainty. This required the definition of scan paths for uncertainty prior to the experiment. In Figure 8, the estimation was conducted using this definition. The benefits of this approach are that it is easy to understand the viewing behavior by observing the eye movements, and the uncertainty may directly affect some typical eye movements. According to this procedure, all scan paths were extracted and determined using the definition, however. If there were few supposed scan paths during observation, this approach was not significant, such as with Subject C in Table V. Therefore, this approach requires well defined scan paths in response to uncertainty.

Another approach to SOB estimation using features of fixations was developed to avoid observation of all scan paths during chosen answer viewing. This approach using SVM may discriminate between levels of SOBs based on the latent features of the fixations. Also, as the SVM can conduct a multiclass discrimination task, this approach will be applied to multiclass SOB estimations in the future. SVM discrimination requires training data in advance, including prototype data consisting of pairs of fixation features and the levels of SOBs. The definition of certainty associated with eye movements is important, as is the first approach using scan paths. Additionally, feature selection issues and understanding of feature space should be considered in this approach. When the training data features did not sufficiently reflect the SOBs, the SVM discrimination failed to estimate SOBs, such as with Subject A. A unique combination of features was required to obtain significant performance for subjects, such as with Subject B in Table V.

While this discussion suggests that careful observation of scan paths or features of fixations should be considered when applying these techniques, individual differences also should be taken into account as in general psychological experiments. As the authors stated in the introduction, certainty estimation using eye movements can be used in a variety of applications that estimate the SOB of users, in particular by people with motor control disabilities. A robust discrimination procedure is required. The previously mentioned challenges of establishing a robust procedure will be topics of our further study.

7. SUMMARY

This article proposes two estimation methods of "strength of belief" (SOB) for answers to multiple choice tasks using eye movements for fixation points, where subjects were looking at answers they had chosen. The high and low SOBs for the answers are defined according to each subject's declaration of certainty. Their performances are evaluated to determine the feasibility of each subject's certainty. The results are summarized as follows:

(1) To produce an SOB estimation procedure using scan paths of eye movements, all fixation points were classified into cells which corresponded with the positions of answers. To detect low SOBs for answers, an estimation model consisting of three cell transactions was created. As a result, the estimation performance with a model using scan-path analysis of eye movements was significant, and the performance was high when the threshold for classifying high and low SOB was set to low. Both precision and recall rates for high SOB were higher than the rates for low SOB.
(2) The SOB estimation procedure using Support Vector Machines with features of fixations was developed without processing scan paths. The estimation performance was also significant. The discrimination performance for high SOB was high, however the performance for low SOB was insufficient.
(3) According to both cases of estimation performance, the results provide evidence that "strength of belief" for answers can be estimated significantly using eye movements.

Both estimation procedures have advantages only for high SOB, but the overall performance is suppressed, however. Therefore the improvement of estimation performance should be addressed.

Also, the development of a more robust estimation method with appropriate features of eye movement will be the subject of our further study.

REFERENCES

Brooke, J. 1996. SUS: A quick and dirty usability scale. In *Usability Evaluation in Industry*, P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, Eds. Taylor & Francis, London, UK, 189–194.

Chang, C. and Lin, C. 2001. *LIBSVM: A Library for Support Vector Machines*. `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Choi, Y. S., Mosely, A. D., and Stark, L. W. 1995. String editing analysis of human visual search. *Optometry Vision Sci. 72,* 7, 439–451.

Google. 2007. Web search help center. `http://www.google.com/support/bin/answer.py? answer=504&topic=532`.

Imai, H. and Takano, Y. 1995. Kioku wo tadoru (Pursuing memory—in Japanese). In *Memory*, Y. Takano, Ed. University of Tokyo Press, Tokyo, Japan, 29–48.

Ishizaki, M. and Den, Y. 2001. *Discorse and Dialogue* (in Japanese). University of Tokyo Press, Tokyo, Japan.

Iwahashi, N. 2003a. Language aquisition by robots—towards new paradigm of language processing. *Trans. Japan. Soc. Artif. Intel. 18,* 1, 49–58.

Iwahashi, N. 2003b. A method for the coupling of belief systems through human-robot language interaction. In *Proceedings of the International Conference on Robot-Human Interaction (ROMAN'03)*. IEEE, 385–390.

Jackson, P. and Moulinier, L. 2002. *Natural Language Processing for Online Applications*. John Benjamins Publishing, The Netherlands.

JSVS (Japan Society of Vision Science), Ed. 2001. *Shikaku jyouhou syori handobukku* (Handbook of Visual Information Processing—in Japanese). Asakura Shoten, Tokyo, Japan.

NAC IMAGE TECHNOLOGY 2007. Eye mark recorder 8. http://www.eyemark.jp.

NAC IMAGE TECHNOLOGY 2004. Eye mark data analysis software, model SP506(Ver. 3.21).

NAKAYAMA, M. AND TAKAHASI, Y. 2006. An estimation of certainty for multiple choice responses using eye-movement. In *Electronic Proceedings of the 2nd Conference on Communication by Gaze Interaction—(COGAIN'06)*. 67–72.

NAKAYAMA, M. AND TAKAHASI, Y. 2007. An estimation of response certainty using eye-movements. In *Proceedings of European Symposium on Artificial Neural Networks*. d-side Publications, 91–96.

NOTON, D. AND STARK, L. 1971. Scanpaths in eye movements during pattern perception. *Science* 171, 308–311.

PALMER, S. 2000. *Vision Science—Photons to Phenomenology*. The MIT Press, Cambridge, MA, USA.

PUOLÄMAKI, Y., SALOJÄRVI, J., SAVIA, E., SIMOLA, J., AND KASKI, S. 2005. Combining eye movements and collaborative filtering for proactive information retrieval. In *Proceedings of the Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval 2005*, A. Heikkil, A. Pietik, and O. Silven, Eds. ACM Press, New York, USA, 145–153.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psych. Bull. 124,* 3, 372–422.

SALOJÄRVI, J., PUOLÄMAKI, K., AND KASKI, S. 2004. Relevance feedback from eye movements for proactve information retrieval. In *Proceedings of the Workshop on Processing Sensory Information for Proactive System (PSIPS 2004)*, A. Heikkil, A. Pietik, and O. Silven, Eds. Oulu, Finland.

SHIGEMASU, K. AND UENO, M. 1993. A new item response model with parameters reflecting state of knowledge. *Behaviormetrika 20,* 2, 161–169.

SHIMODAIRA, H., NOMA, K., NAKAI, M., AND SAGAYAMA, S. 2001. Dynamic time-alignment kernel in support vector machine. In *Advances in Neural Information Processing Systems 14, Vol. 2*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. The MIT Press, Cambridge, MA, 921–928.

SHINOTSUKA, H. 1993. Confidence and accuracy of mental judgments. *The Japanese J. Psych. 63,* 6, 396–403.

STORK, D. G. R., DUDA, O., AND HART, P. E. 2001. *Pattern Classification*, 2 Ed. John Wiley & Sons, Inc. Japanese translation by M. Onoue, New Technology Communications Co, Ltd, Tokyo, Japan.

TOKUNAGA, T. 1999. *Information Retrieval and Natural Language Processing* (in Japanese). University of Tokyo Press, Tokyo, Japan.

TSUDA, K. 2003. Ka-neru hou no riron to jissai (Theory and practice of kernel method—in Japanese). In *Pattern ninsiki to gakusyuu no toukeigaku (Pattern Recognition and Statistics of Learning)*, S. A. et al., Ed. Iwanami Shoten, Tokyo, Japan.

UNDERWOOD, G. 2005. Eye fixations on pictures of natural scenes: Getting the gist and identifying the components. In *Cognitive Processes in Eye Guidance*, G. Underwood, Ed. Oxford University Press, Oxford, UK.

ZHANG, Y. 2007. *An Item Response Model on Probability-Testing* (in Japanese). University of Tokyo Press, Tokyo, Japan.

Received April 2007; revised August 2007, December 2007; accepted December 2007