Development of a Web Application: Recording Learners' Mouse Trajectories and Retrieving Their Study Logs to Identify the Occurrence of Hesitation in Solving Word-Reordering Problems

Mitsumasa ZUSHI
Graduate School of Informatics, Shizuoka University
Hamamatsu, Shizuoka Prefecture 432-8011, Japan

and

Yoshinori MIYAZAKI
Graduate School of Informatics, Shizuoka University
Hamamatsu, Shizuoka Prefecture 432-8011, Japan

and

Ryosuke MIKI
Graduate School of Informatics, Shizuoka University
Hamamatsu, Shizuoka Prefecture 432-8011, Japan

and

Ken NORIZUKI
School of Information Studies, Shizuoka Sangyo University
Fujieda, Shizuoka Prefecture 426-8668, Japan

ABSTRACT

Most computer marking systems evaluate the results of the answers reached by learners without looking into the process by which the answers are produced, which will be insufficient to ascertain learners' understanding level because correct answers may well include lucky hunches, namely accidentally correct but not confident answers. In order to differentiate these lucky answers from confident correct ones, we have developed a Web application that can record mouse trajectories during the performance of tasks. Mathematical analyses of these trajectories have revealed that some parameters for mouse movements can be useful indicators to identify the occurrence of hesitation resulting from lack of knowledge or confidence in solving problems.

Keywords: e-learning, Mouse Trajectory, Study Logs, Information Retrieving Tool, Occurrence of Hesitation.

1. INTRODUCTION

Based on the supposition that answers being correct or incorrect alone may not be a true reflection of the learners' understanding unless the processes of the learners' responses are carefully considered, we have been developing a Web application which will enable both teachers and learners to notice the crucial aspects of "uncertainty" or "hesitation" in producing answers resulting from learners' lack of confidence or knowledge. Through the analyses of our former experiments, where learners were asked to solve word-reordering problems (WRPs) by dragging and dropping the given words with the mouse, we have verified the possibility for several parameters for mouse movements showing hesitation or uncertainty, to detect the difficult problems for a group of learners as a whole.

In this paper, we first describe our Web application briefly, and then find the norms of mouse trajectories for correct and confident answers, and compare the norms with mouse trajectories included in the answers of which learners are not certain or confident, in order to investigate whether or not there are any significant differences between them. Since confident and correct answers will show smooth mouse movements in placing the words in the right position, these differences will be the clues for detecting unnecessary movements in answering WRPs, which, we hypothesize, will deeply connect with learners' uncertainty and lack of knowledge or confidence. We also suggest probable ways to apply the "differences" from the norms to identify hesitation or uncertainty in individual learners' solving process of each problem in order to detect which part or words are difficult for a particular learner to reorder and produce a correct answer.

2. RELATED RESEARCH

There have been several studies dealing with study logs or mouse movements so that teachers or administrators would be able to know users responses during their task performance. In order to develop an authentication system, Tateda, et al. illustrated a number of technical properties of learners' mouse movements [1]. Ohmori, et al. investigated mouse behavior in the course of learners' reading task, classifying learners' reading habits into three patterns [2]. Arroyo, et al. created a web logging system that helps website administrators check usability and analyze the collected data [3]. Ikegami scrutinized the learners' study habits from the logs and suggested effective learning environment [4], [5]. Developing their own software package called "MouseTracker," Freeman and Ambady enabled teachers to monitor subjects' behaviors during a psychological task, with the mouse trajectories being visualized in real time [6]. Nakamura, et al. analyzed facial movements in order to know the difficulty/easiness of the e-learning problems the learners solved [7], and Horiguchi, et al. developed a system that can presume learners' state of mind during the course of e-learning tasks by analyzing their facial responses and mouse speed [8].

Though featuring the analysis of study log and mouse movements, these studies have not analyzed the processes of learners' solving problems, as discussed in Miyazaki, et al. [9] and Zushi, et al. [10], which, the present authors believe, are the potential clues to ascertain learners' understanding levels.

3. WEB APPLICATION

In order to achieve our goal, we have been developing software that has three modules with independent, but also interrelated functions: 1) Study Module, which requires learners to perform word-reordering tasks by "dragging and dropping" each word into the appropriate position in a sentence, simultaneously recording all the mouse trajectories as well as the timing of drag-and-drops (D&Ds) in answering the problems; 2) Problem Construction Module, where teachers construct and add new problems on their own, or change marking scales from the default form; and 3) Retrieval & Analysis Module, which reproduces all the actions recorded in the learners' mouse trajectories, and analyzes the data from the diverse patterns of the study logs both from the learners' and problems' perspectives. This software, consisting of three modules, has been programmed using PHP, Visual Basic, and MySQL.

Study Module

In WRPs, learners are required to make an English sentence from given words, one with a meaning equivalent to the sentence provided in the learners' native language. This type of problem has been and is even now a popular means in Japan to measure learners' knowledge of grammatical items, sentence structure, idioms/idiomatic phrases, and usages--the command of which is essential for producing correct sentences. Since Japanese is a synthetic language, it has quite a loose word order, while in an analytic language like English, word ordering has a crucial importance to decide the meaning of the sentence, its grammaticality, and acceptability. This is why Japanese novice or poor learners of English have difficulty in making a correct English sentence, and WRPs are effective in confirming the ability needed to produce English sentences.

The words to be rearranged are given in the "problem slot," and all the words should be moved into the "answer slot" by D&Ds with the mouse. All mouse movements (click, drag, drop, and their timing) for solving problems are recorded here (see Figure 1).

It is required for learners to press the OK button to finish answering, and then they rate their confidence in the answers on a pull-down, four point scale. After all these procedures are completed, the answers are automatically evaluated by the system.

In order to facilitate learners' performing the tasks, this module has the following functions:

1) Word groupings: an arbitrary number of words can be grouped together by mouse-dragging (rectangular selection) if it is convenient for learners to treat them together, and
2) Relocation to registers: areas called "registers" are provided as a temporary "shelter" for words, where a set of words can be integrated into meaningful segments if it is preferred for learners to organize their ideas.

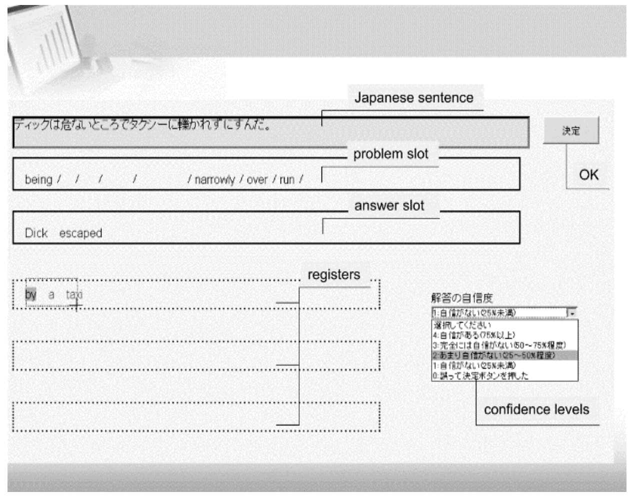

Figure 1. Study Module

Problem Construction Module

This module helps teachers provide WRPs for learners to solve. Figure 2 illustrates the processes of making problems briefly.

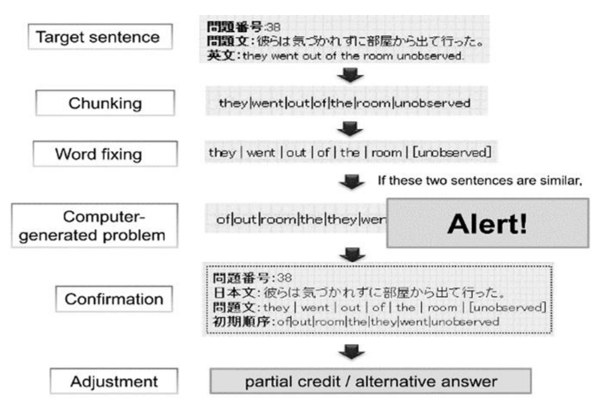

Figure 2. Procedure for a WRP construction

All teachers have to do is just to type correct sentences in English and the corresponding meaning in Japanese. Teachers may use "chunking" and/or "word fixing" functions if they want to make the problem easier or shorter. By using the chunking function, a series of words become one group when the straight-bar(s) "|" that separate each adjacent words are deleted, and the group of words will be treated as one word, making the number of words included in the problem smaller. The word fixing function is also available when teachers prefer to make the problem easier or shorter by literally fixing the word position in a sentence. Once a word is fixed to a certain position, learners are not able to change it, which means that learners have only to move the rest of the words in the problem.

Furthermore, this function is the most useful when teachers wish to avoid double/multiple correct answers to one problem. The computer's binary recognition system accepts only one generated sentence as the correct answer unless another or other correct ones have been additionally programmed in advance. The "word fixing" function makes it easier to deal with a sentence composed of words that can be arranged into more than one correct order. For example, the sentence, "They went out of the room unobserved" can also be ordered correctly as "Unobserved they went out of the room." Using this function, though, teachers can fix the position of "unobserved" in the initial or end position (as in Figure 2), and thereby prevent the possibility of double correct answers.

After these procedures have been through, teachers can choose one of three break down methods: optional, alphabetical, or randomized order (Figure 3).

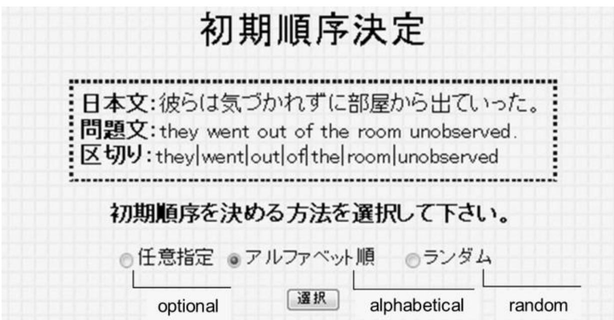

Figure 3. Choice of a break down method

Automatic randomization is the easiest to use and will be the most preferred, but there is a slight, yet rather serious danger of producing a WRP which is similar in form to the correct sentence, especially when the number of words to be moved is as small as seven or eight. The similarity to the correct sentence will make the problem much easier for learners to solve, thus fewer movements of the mouse will be necessary in solving it. If automatic randomization produces a similar word order to the target structure, the teacher can repeat using randomization until the word order in the problem is quite different from that of the correct sentence. Another break down method is an alphabetical word order, in which chances for the similarity to happen become smaller. However, alphabetical-order arranging may not be perfect, either, creating a problem that has already formed an important structure used in the sentence. An "Alert" function is incorporated into this module to eliminate a teacher's burden of checking these similarities between the problems' word orders and correct sentences. When notified by "Alert" that the problem created has a similar word order to the answer, namely, the pre-specified ratio of words--or more--are already in order, or a crucial idiom, which might be worth some partial credit, is already in the correct order, teachers will never fail to change the word order in the problem by randomizing again or using a third method, the optional word order, which allows the teacher to choose the order.

In the last stage of making problems, confirming the target sentence (=answer) and the word arrangements in the problem, teachers can prepare partial credit for a correct part of a sentence or a sentence which is not completely correct but good enough to make sense (e.g. Olympics are watched by the people all over the world).

Retrieval & Analysis Module

The study log data recorded by the Study Module when learners perform tasks, can be retrieved and analyzed by using one of the six menus in the Retrieval & Analysis Module. The menus included in this module are Reproduction, Learner Analysis, Problem Analysis, Study Log Retrieval, Correlation Analysis, and Clustering menus, each of which has links to related pages of other menus so that teachers are able to have an access to sets of data they want to check.

1) Reproduction: mouse trajectories are reproduced in visible lines whose color changes in every ten seconds so that the newer lines can be identifiable when the trajectory lines overlap, with reproduction speed changeable from 50 % to 500% of its original speed, and also with reproduction being able to start at any selected point on the slider (Figure 4).

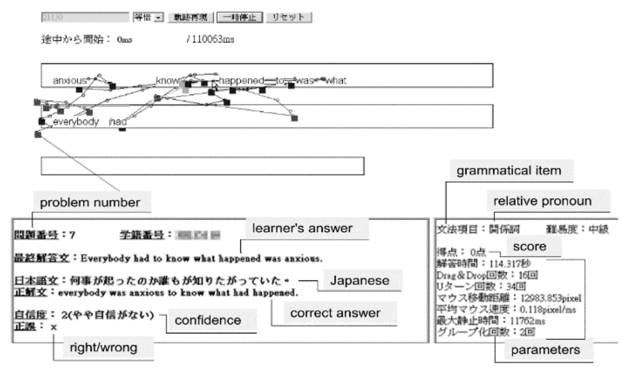

Figure 4. Reproduction of mouse trajectories

2) Learner Analysis: data for a particular learner is available, such as the percentage of correct answers classified by grammatical items, the total number of problems attempted, the total elapsed logon time, the percentage of overall correct answers, as well as the average time needed for answering problems (Figure 5).

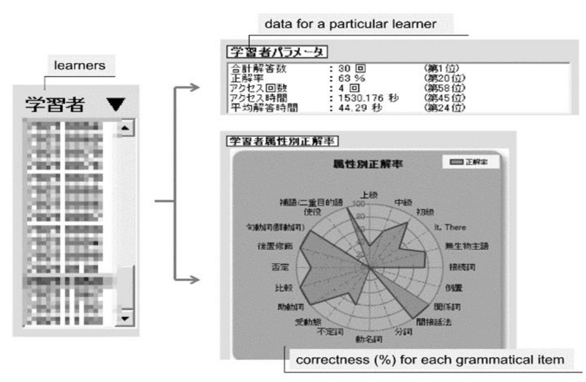

Figure 5. Learner analysis

3) Problem Analysis: data for a particular problem can be reviewed, such as the number of times the problem was attempted, the percentage of correct answers, the average time needed for answering, and so on (Figure 6).

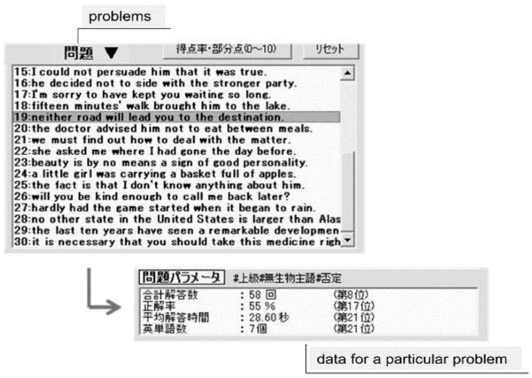

Figure 6. Problem analysis

4) Study Log Retrieval: this menu helps teachers retrieve data for specific criteria they want to focus on. Combining several search types and commands, the teacher can easily have an access to, for example, learners that used U-turns --the right-and-left, or up-and-down mouse movements--more than 15 times to solve a particular problem, learners that have used this software in the last two months, or problems that required D&Ds less than 10 times for a certain learner to solve, and so on.
5) Correlation Analysis: with this menu correlation analysis is available for all combination of parameters showing the elements of the mouse movements--such as response time needed, the number of D&Ds, the total distance of mouse movement, and the number of U-turns--from either the learners or the problems perspective. Figure 7 is an example of correlation between the average time for answering and the average scores, with each dot representing a particular learner. We see a roughly negative correlation here, which means that the less time needed for answering, the higher the scores.

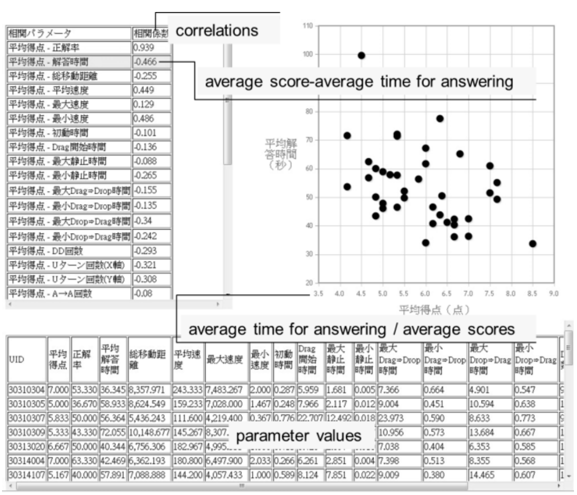

Figure 7. Correlation analysis (the average time for answering/the average scores)

6) Clustering: this function assembles similar learners in groups. In this study, considering the large difference in the variances and scales of each parameter, a standardized Euclidean distance was used instead of the generally used (non-standardized) Euclidean distance. Ward's method was adopted for computing the distances among clusters. Teachers can choose the number of clusters and which parameters to be incorporated for clustering criteria, from "the number of problem attempted per learner," "the ratio of correct answers," "the number of times accessed," "total elapsed logon time," and "average answer time."

The newly incorporated function for the latest experiment to be conducted is a function to retrieve the trajectories of how each word is treated by each D&D. The former system counted the total number of D&Ds, U-turns, and the time needed in solving a whole problem, but the new system enables us to observe each word movement. D&D(s), U-turn(s), and the time used in the treatment of one word, as well as the time elapsed between a particular drop and the click of the next word (D-C time), namely the time between a particular D&D and the next D&D, and standstill time of the mouse can be retrieved now. These data are expected to contribute to reveal more delicate points in the course of performing tasks.

4. EXPERIMENT and DISCUSSION

Asking 40 students with a variety of majors at a certain university in Japan, to solve 30 problems using the e-learning system we have developed, we conducted an experiment in which the subjects were informed that their mouse trajectories would be recorded and analyzed for the purpose of investigating the solving processes. 1,178 sets of data were collected in total, with 22 sets of data being inappropriate for analysis owing to a recoding failure.

Confident and Correct Answers as Norm

The answers the subjects reached in our former experiment fall into one of four categories: 1) confident and correct, 2) confident but not correct, 3) not confident but correct, and 4) not confident and incorrect. Confident answers belonging to categories 1 and 2 include useful mouse movement logs for finding the norm of "answers without hesitation or uncertainty," although the learners who produced category 2 type of answers should review and learn the correct sentence. The answers from category 3 need very careful treatment because most computer marking systems may evaluate them as "passing" in spite of the fact that learners are not yet sure of the correct answer, or they might have reached the correct answers accidentally. There is no need of analyzing the answers of category 4 for the purpose of ascertaining learners' understanding levels because they clearly do not understand, which means that they have to review from the beginning, but these answers also offer very useful data for uncertainty or lack of confidence in terms of mouse trajectories.

The answers that fell into category 1 totaled 207 in the latest experiment, which we use as the norms, and compare the norms with mouse trajectories included in 171 answers of category 4 (in which learners are not certain or confident), in order to investigate whether or not there are any significant differences between them. Since confident and correct answers will show smooth mouse movements with the least hesitation, these differences will lead to the clues for detecting unnecessary and uncertain movements caused by hesitation owing to lack of confidence or knowledge.

Comparison between Category 1 (Confident and Correct) and 4 (Not confident and Incorrect)

Focusing on several parameters for mouse movements, we compare the distributions of the answers in two categories. The following graph is a comparison of time needed for answering each problem (Figure 8).

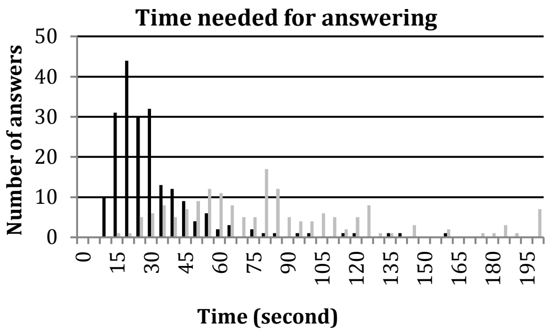

Figure 8. Distribution of time needed in the two types of answers: black-category 1, and gray-category 4

The difference is clearly seen in the graph, but in order to numerically illustrate the degree of difference between the two types of answers, we calculate the deviation: when the norm is determined on the basis of category 1, deviation calculated in terms of the average for the answers in category 4 shows 97.36, with 84.2% of answers of category 1 being higher than 65 in deviation value.

Another comparison is of the longest D-C time during performing tasks, as is seen in Figure 9 below.

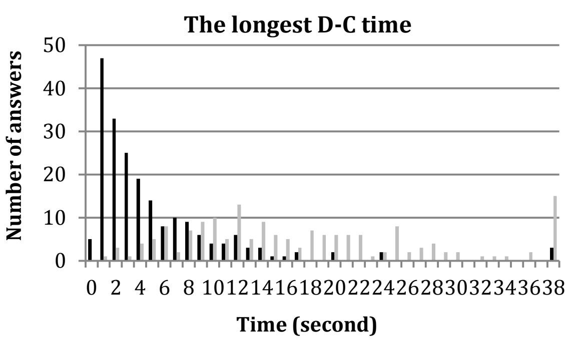

Figure 9. Distribution of the longest D-C time in the two types of answers: black-category 1, and gray-category 4

The deviation value here is 81.73, with 75.44% of answers showing higher values than 65.

Calculating deviations by comparing the average values from the answers of one group (category 4) with the norms determined on those of another group (category 1), in a strict statistical sense, might not be the most appropriate way, but it will suffice to demonstrate that two groups are totally different or almost the same.

Other comparisons (average velocity of mouse movement, the longest standstill time, and the number of U-turns) also reveal pretty large differences between the two types of answers, while time elapsed from completion of reordering to pushing OK button (which is expected to be the time for confirmation) indicates no significant difference between the two groups (Table 1).

Table 1. Deviation of parameters in category 4 (not confident and incorrect answers) from the norm

	P1	P2	P3	P4
Deviation	35.91	72.95	V:73.59 H:69.40	49.65

P1: Average velocity of mouse speed
P2: The longest standstill time
P3: The number of vertical (V) and horizontal (H) U-turns
P4: Time elapsed from completion of reordering to pushing OK button

Distribution of Category 3 (Not Confident but Correct) Type of Answers

The parameters showing large differences from the norms, and indicating lack of confidence or knowledge required in solving the problem, can be included even in correct answers, as many cases found in 72 answers of category 3 (not confident but correct). Time for answering in category 3 (correct answers) and 4 (incorrect answers) in Figure 10 does not illustrate a large difference between the two as clearly seen in Figure 8, which also compares the time for answering in correct answers and incorrect answers.

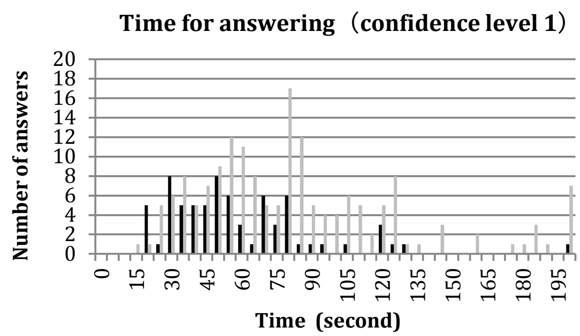

Figure 10. Distribution of time needed in the two types of answers: black-category 3, and gray-category 4

Similarly, the longest D-C time in category 3 and 4 shows little difference as in Figure 11.

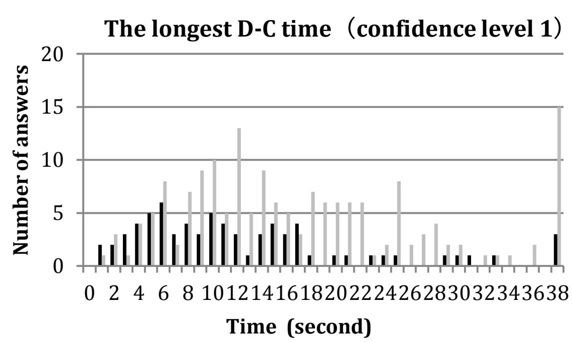

Figure 11. Distribution of the longest D-C time in the two types of answers: black-category 3, and gray-category 4

Although the answers belonging to category 3 are correct and those of 4 are incorrect, the mouse trajectories of the answers both in categories 3 and 4 are nevertheless similar in terms of the parameters for time needed for answering and D-C time. The learners whose answers include these trajectories with parameters evaluated as "not confident," need to review the same type of problems because they have not yet reached an adequate understanding of the target grammatical items, such as sentence structure, collocations, usages, or idioms irrespective of the correctness of their answers.

5. CONCLUDING REMARKS

The results of our experiment and analyses reasonably suggest that the parameters showing large differences from the norms, (especially time needed for answering, D-C time, average velocity, standstill time, and the number of U-turns), can be used as clues deeply connected to lack of confidence and hesitation of mouse movements, and these parameters will be applicable to identify hesitative part(s) in individual learners' solving process of each problem when a proper mathematical treatment is given to mouse trajectories produced by individual learners.

Though insufficient in number to illustrate the distribution or to calculate correlation coefficients with confidence levels, the adjusted number of D&Ds (the number of words included in the problem is reduced from the total number of D&Ds in order only to calculate the pure number of D&Ds that exceeded the minimum number necessary), the D&Ds of the same word, and the number of D&Ds in the answer slot (replacements of words in the answer slot) are also assumed to be indicative of hesitation in the process of performing tasks since they are excessive mouse movements. They are very likely to be useful as clues to locate where in the process of solving one problem hesitation occurs.

In order to verify that these D&Ds clearly feature difficulties the learners are experiencing and in order to ascertain accurately when and where hesitations occur in their positioning of the words in the target sentence, our future tasks should include:

1) finding appropriate ways to treat the D&Ds' data mathematically for the purpose of connecting them to learners' understanding levels,
2) calculating personal norms of each learner to measure the difference of the important parameters showing hesitation,
3) incorporating the automatic retrieval function of the important parameters in the process of solving one problem to determine where in each reordering the hesitation occurs, so that they will be displayed on the screen of learner analysis menu, and
4) combining the parameters in some pattern to examine how they are useful for more accurate identification of hesitation.

To collect adequate data for important parameters which will contribute to personal analyses, we need to conduct experiments in which learners are required to solve more than 30 problems as well as to continue improving the system.

ACKNOWLEDGMENT

This work was supported by a Grant-in-Aid (KAKENHI) for "Scientific Research (C) (25370621)."

REFERENCES

[1] C. Tateda, et al., "Evaluation of an Individual Authentication System Using the Trajectory by the Mouse", **IEICE Technical Report, HCS** 105(74), 2005, pp. 17-20.
[2] Y. Ohmori, et al., "Understanding Learning Effect Using MBC (1)-Analysis of Mouse Behavior", **Conference on IT Education**, 2006, pp. 51-54.
[3] E. Arroyo, et al., "Usability Tool for Analysis of Web Designs Using Mouse Tracks", **Proceedings of CHI EA'06, ACM**, 2006, pp. 484-489.
[4] M. Ikegami, "Enhancing Students' English Reading and Listening Skills through the Use of the Computer-assisted Intensive English Training Program", **Journal of Language and Literature**, Vol. 26, No. 1, 2006, pp. 103-26.
[5] M. Ikegami, "Enhancing Student's English Reading and Listening Skills through the Use of the Computer-assisted Intensive English Training Program III--The Effect of Learning Environment and Course Design--", **Journal of Language and Literature**, Vol. 29, No. 1, 2009, pp. 229-257.
[6] J. B. Freeman, and N. Ambady, "MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method", **Behavior Research Methods**, Vol. 42, No. 1, pp. 2010, 226-241.
[7] K. Nakamura, et al., "Estimating Learners' Subjective Impressions of the Difficulty of Course Materials by Observing Their Faces in E-Learning", **Journal of the Institute of Electronics, Information and Communication Engineers**, Vol. J93-D, No.5, 2010, pp. 568-578.
[8] Y. Horiguchi, et al., "An Automatic Estimation System of Learners' Impasses Based on Features in Low-Level Interactions in an E-Learning Environment", **The 24th Annual Conference of the Japanese Society for Artificial Intelligence**, 2010, 2F1-3.
[9] Y. Miyazaki, et al., "Web Application for Recording Learners' Mouse Trajectories and Retrieving Their Study Logs for Data Analysis", in eds. J. Cheng, G. Fu, and X. Han, **Technology Enhanced Language Learning**, 2011, pp. 95-99.
[10] M. Zushi, et al., "Web Application for Recording Learners' Mouse Trajectories and Retrieving their Study Logs for Data Analysis", **APTEL 2011 Best papers: Knowledge Management & E-Learning: An International Journal**, Vol. 4, No. 1, 2012, pp. 37-50.