A study on Extraction Method of Pairs of Questions With Opposite Meanings Based on Distance between Questions

Yosuke Watanabe*, Tomohiro Yoshikawa†, Takeshi Furuhashi‡, Eiji Hirao§ and Hiroto Iguchi§
*Nagoya University
Email: yosuke@cmplx.cse.nagoya-u.ac.jp
†Nagoya University
Email: yoshikawa@cmplx.cse.nagoya-u.ac.jp
‡Nagoya University
Email: furuhashi@cmplx.cse.nagoya-u.ac.jp
§NEC Corporation

***Abstract*—Companies often carry out questionnaire(s) in order to design marketing strategies or to grasp the market trends. Recently, Web questionnaire survey becomes popular with the spread of the Internet in order for companies to reduce cost and to get a lot of questionnaire data. However, in the Web questionnaire survey, some respondents do not answer the questions seriously, because it is not done face-to-face and is done just for giveaway basically. The authors have tried to quantify the conflict of response based on the contradiction represented by the pair(s) of questions with opposite meaning. One of the most important matters in the quantification of the conflict of response is the selection of the pair(s) of opposing questions, which are actually regarded as opposite meanings by respondents. This paper defines the distance between questions which shows the similarity of questions for respondents and proposes a method to derive the pair(s) of opposing questions based on the distance between questions. This paper applies the proposed method to an actual questionnaire data and shows more appropriate opposing questions for respondents can be extracted than the prepared pairs of questions.**

I. Introduction

Companies often employ questionnaire(s), which is the quantified data of respondents' Kansei, in order to design marketing strategies or to grasp the market trends. For example, when companies plan a new project, they often survey the impression of people by a questionnaire. The analysis of a questionnaire data gives some marking strategy with the prediction of marketing scale or the target groups to sell or some hints for the modification of products [1][2]. The rating scale method is widely used in questionnaires [3][4], designed to obtain impressions for evaluation objects such as products, services and brands. This method requires plural evaluation objects and questions, and respondents answer the sets of questions for each evaluation object according to their impressions with multiple grading scales. In this way, people's impressions on evaluation objects can be quantified as the graded data [5].

Questionnaires data used to be carried out by face-to-face survey or mail-in survey [6]. Recently, Web questionnaire survey has become popular with the spread of the Internet in order for companies to reduce the cost and to get a lot of questionnaire data in a short time. However, in the Web questionnaire survey, some respondents do not respond the questions seriously; for example, they might respond to them randomly or equally without reading questions, because it is not done face-to-face and is basically done to earn giveaway or points. If those responses are included into questionnaire data, there is a possibility that the analysis result of the data is not accurate and shows different characteristics or trends.

The authors have proposed a quantification of the conflict of response and the respondent's interest based on the evaluation values [7][8]. The conflict of response is based on the contradiction by the pair(s) of questions with opposite meaning. The respondent's interest is defined as the average of variances of evaluation values for each evaluation object. One of the most important matters in the quantification of the conflict of response is the selection of the pair(s) of opposing questions, which are actually regarded as opposite meanings by respondents. In conventional approach, the authors have prepared the pair(s) of questions with opposite meaning in the design of questions. However, it has not been sure whether the respondents actually regarded these questions as the opposite meaning in the questionnaire, and it might be better pair(s) in the questions. It is necessary to evaluate the actual impression of questions for the respondents.

This paper defines the distance between questions [7] which shows the similarity of questions for respondents and proposes a method to derive the pair(s) of opposing questions based on them. The distance between questions enables us to verify whether the prepared pairs of opposing questions were actually regarded as opposite meanings for respondents and may let us find more appropriate pair(s) of questions than the prepared ones.

This paper applies the proposed method to an actual questionnaire data. It shows that the prepared pairs of opposing questions were not necessarily regarded as opposite meaning

for the respondents and more appropriate pairs for the quantification of the conflict of response can be extracted, which is investigated through the comparison of the average character counts in free text form.

II. Proposed Method

A. Conflict of response

Conflict of response [7][8] is quantified using pair(s) of questions which have opposite meaning each other. In the proposed method, the contradiction of response, i.e., evaluation values, is defined as quasi-coefficient of correlation between opposing questions that employs the median value of grading scales instead of the average of evaluation values. For the graded data obtained through a rating scale method, evaluated values for the paired questions, 1 and 2, on i-th object are defined as x_{i1} and x_{i2}, and the median value is defined as M. The value of conflict of response C is defined as follows:

$$C = \frac{\sum_{i=1}^{n}(x_{i1} - M)(x_{i2} - M)}{\sqrt{\sum_{i=1}^{n}(x_{i1} - M)^2}\sqrt{\sum_{i=1}^{n}(x_{i2} - M)}^2} \quad (1)$$

where n denotes the number of evaluation objects. C runs from -1 through 1, and C becomes large when the evaluated values for the paired questions with opposite meanings are both lower or higher than the M. Thus, if a respondent's C is large, the data of the respondent concerns that there is a possibility to have conflict in his/her response, because he/she might have answered without reading questions.

Note that it is not effective to prepare the questions such as "cool" and "not cool," because the expression is so obvious that the respondents become aware of the intent. Moreover, the answer of these two questions gives only one response. Reduction of the number of questions is one of the most important issues in questionnaires because of humans' fatigue for answering. Then the loss of information above causes inefficient questionnaire especially in the rating scale method in which the respondents answer to the same set of questions to the number of evaluation objects.

B. Derivation of pairs of opposing questions based on distance between questions

1) Distance between questions: The distance between questions [8] is defined as a distance between the evaluated values of two questions in the original space generated by the axes of evaluation objects. It is assumed that the distance between questions in the evaluation object space represents the similarity of them for a respondent. When the evaluated values of question a and b for evaluation object i ($i \in 1...N_o$, N_o: number of objects) in respondent k ($k \in 1...N_r$, N_r: number of respondents) are defined as x_{kia} and x_{kib}, respectively, the distance between the questions, d_{ab}, is defined as follows:

$$d_{ab} = \frac{1}{N_r}\sum_{k=1}^{N_r}\sqrt{\sum_{i=1}^{N_o}(x_{kia} - x_{kib})^2} \quad (2)$$

When the distance d_{ab} is small, it means that the respondents gave similar evaluation values to plural evaluation objects, and it can be thought that the respondents regarded the question a and b as similar meaning. (For example, when the respondents evaluated in the question a, "it feels cool," to the evaluation object A and B as (A, B) = (5, 2), and question b, "it feels good," as (A, B) = (4, 2), the evaluated value to the question a was synchronized with b for the different evaluation objects, then "cool" and "good" for them could be similar meaning each other in the questionnaire.) When the distance d is large, it can be said that they regarded these questions as different meaning. Relationships among questions based on their distances can be visualized using a dendrogram shown in Fig.1 [8].

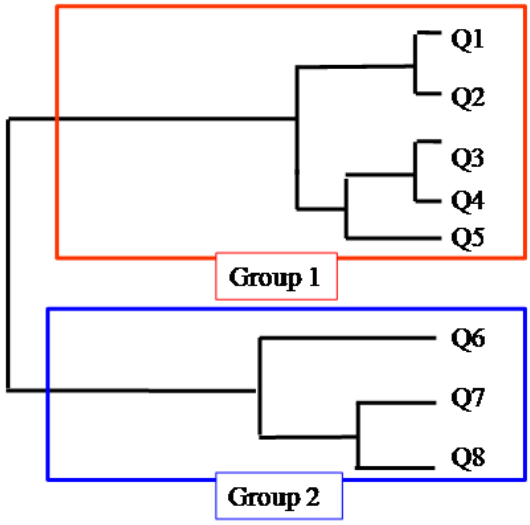


Fig. 1. Example of visualization by dendrogram

2) Derivation of candidates of pairs of opposing questions: It is expected that the ideal pair of opposing questions have large distance d because they are opposite meaning each other, and then d becomes small when evaluated values of either of them are reversed, for example, "1" is exchanged for "5" and "4" for "2." The proposed method derives the candidates of pairs of opposing questions in the following steps. First, the questions are visualized based on their distances using a dendrogram like Fig.1. It is expected that the questions are clustered into some groups in which they have similar tendency of evaluated values, e.g., group 1 and 2 in Fig.1. Next, the evaluated values in one of these groups are reversed and the distance between questions is re-visualized as shown in Fig.2, which is an example of the reverse of evaluation values in Group 2 in Fig.1. In Fig.2, Group 1' has Q7 which was in Group 2 before the evaluated values were reversed while Q6 and Q8 are still far from the questions in Group 1'. Finally, it derives Q1-Q7, Q2-Q7, Q3-Q7, Q4-Q7, and Q5-Q7, which are satisfying the assumption for the ideal pair described above, as the candidates of the pairs of opposing questions. On the other hand, the questions in Group 2' which keep the

distance between questions large in this reverse operation are not selected as the candidates, because they were regarded as different meaning from others but were not regarded as opposite meaning by the respondents.

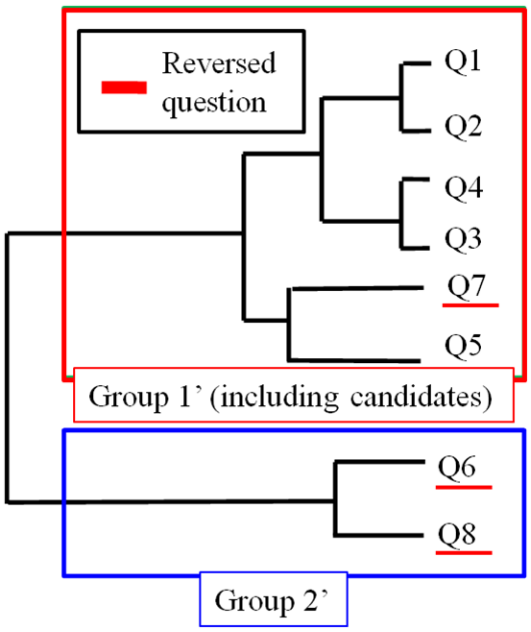


Fig. 2. Dendrogram after reverse of evaluation values

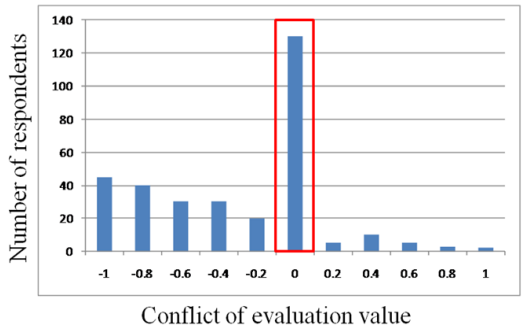
3) Derivation of pairs of opposing questions: The pair(s) of opposing questions are finally derived from the candidates in II-B2. First, the conflict of response C for each respondent is calculated by eq.(1) and visualized by histogram, as shown in Fig.3, for every candidate. Next, the histogram for each candidate is compared based on the following ratios in all respondents:

(a): the ratio of respondents to be $C = 0$

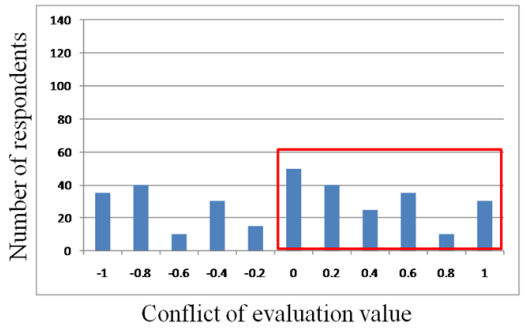
(b): the ratio of respondents to be $C > 0$

Fig.3(c) shows an example of the histogram having an ideal feature. If the ratio of respondents to be $C = 0$ is high as shown in Fig.3(a), it is thought that many respondents gave evaluated value M, median value of grading scales and it usually means "neither of applicable nor non-applicable," for either of the pair of questions (see eq.(1)). Then the question would be ambiguous or difficult to answer clearly, and this pair of questions is not appropriate for using in the quantification of the conflict of response. On the other hand, if the ratio of respondents to be $C > 0$ is high, as shown in Fig.3(b), the pair of questions was not regarded as opposite meaning for the respondents.

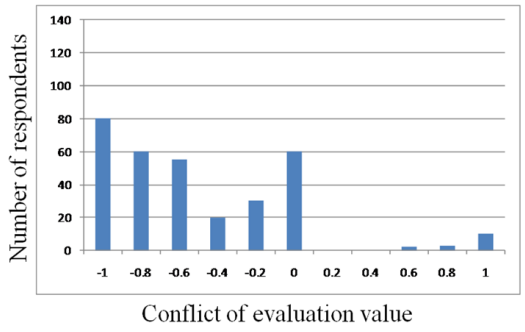
In addition, when a question is overlapped in the candidates as II-B2, only one of them is selected based on these criteria (a), (b). In the case of II-B2, only one pair of opposing questions is selected because Q7 is overlapping in all candidates. Though there should be appropriate thresholds for (a) and (b) to derive the pair(s) of opposing questions, this paper employs relative comparison among the candidates because it is difficult to derive appropriate ones for general data.



(a) Too many respondents with $C = 0$



(b) Too many respondents with $C > 0$



(c) Applicable feature

Fig. 3. Examples of histogram

III. Experiment and Discussion

A. Experimental setting

This experiment of a questionnaire involved 400 respondents, and they answered 15 questions to each of 3 evaluation objects on visualization services for next generation. The questionnaire employed the rating scale method and the respondents were asked to choose one of the five grades 1, 2,

TABLE I
EVALUATION OBJECTS

Obj1	Visualization service for searching: "Pay Service to show you the result of searching information which will be worthful for you."
Obj2	Visualization for after-the-sale service: "Free after-the-sale Service to reply how to handle the trouble in written form to your inquiry."
Obj3	Visualization service for environmental information: "Service to show you environmental performance by the label of authorized energy-saving."

TABLE II
QUESTIONS

Q1	This service will be valued by only a few people.
Q2	It is expensive for the service.
Q3	The object has much disadvantage.
Q4	I want this kind of service.
Q5	I want to use this service if it becomes common.
Q6	This service has a reality.
Q7	It looks trouble in the procedure or the operation.
Q8	This service will contribute the development in business.
Q9	I can not understand because of lack of information.
Q10	I have a special feeling with this kind of service.
Q11	I know this kind of service well.
Q12	I was often asked about this kind of service.
Q13	This service is not necessary.
Q14	I do not want to use this service because it is similar to other existing service.
Q15	This service has high potential for needs.

3, 4, 5 in response to each question. In this survey, grade 5 means "applicable to me" while grade I means "not applicable to me." Table I shows the 3 visualization services used as the evaluation objects, and Table II shows the 15 questions. In addition, the respondents were asked the image of each evaluation object in free-form text. In Table II, Q9-Q11 and Q4-Q13 were the prepared pairs of questions with opposite meaning to calculate the conflict of response in eq.(1).

B. Visualization for relationships between questions

First, distance between questions is calculated, and Fig.4 shows the visualized result by a dendrogram shown in II-B1. Horizontal axis in Fig.4 means the value of distance between questions in the original space. Questions are clustered into 2 groups as shown in Fig.4. It is interesting result that the questions in Group 1 had negative meaning, e.g., Q1 "This service will be valued by only a few people," Q2 "It is expensive for the service" and so on, and vice versa. The meaning of questions for the respondents mainly depended on whether they had positive meaning or not in this questionnaire, which was visualized by the distance between questions defined in this paper.

Next, evaluation values in Group 1 were reversed in accordance with the flow in II-B2, and the distance between questions was re-calculated and re-visualized as shown in Fig.5. In Fig.5, questions are clustered into 3 groups. In

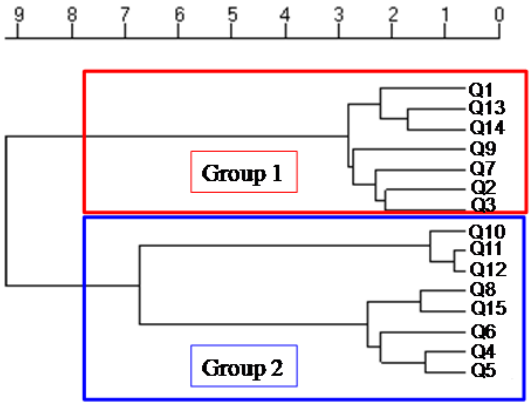


Fig. 4. Distance between questions visualized by dendrogram (before reverse)

comparison with Fig.4 and Fig.5, the questions in Group 1' had similar relationship to Group 1, and they are still far from other questions even if their evaluation values were reversed. It leads the following supposition; though the questions in Group 1' had originally different tendency in the meaning for the respondents comparing with the other questions, the difference of positive/negative meaning in the questions was more emphasized, which made the difference between the questions in Group 1' and the others in Group1 smaller . It is thought that Q3 also had similar reason while it was comparatively far from group 1' as shown in Fig.5. In addition, Q10, Q11, Q12 were far from the others similarly to Fig.5 even when the evaluation values of them were reversed. It is thought they were regarded as different meaning for respondents.

On the other hand, Group 3 in Fig.5 has the questions from both Group 1 and Group 2, which were positive and negative meanings. Then the combinations of (Q1, Q13, Q14) and (Q4, Q5, Q6, Q8, Q15) were derived as the candidates of the pairs of opposing questions (3*5 = 15 candidates) described in II-B2.

Note that the prepared pair of Q9-Q11 had large value of distance between questions both Fig.4 and Fig.5. It is thought that the respondents did not regard these questions as opposite meaning though the questioner had thought they were.

C. Decision of pairs of opposing questions

Histograms of the conflict of response were made on each candidate in III-B and compared their figures. Table III shows the ratio of respondents to be (a) $C = 0$ and (b) $C > 0$ described in II-B3. In Table III, pairs of Q1-Q4 and Q6-Q13 have small values in both (a) and (b). Fig.6 shows the histograms of the conflict of response on Q1-Q4, Q6-Q13, which are described above, and Q4-Q13, which is the prepared one.

In Fig.6, the number of respondents with $C = 0$ in Q4-Q13, the prepared pair, is larger comparing with that in Q1-4 and Q6-13, so it can be said that Q1-Q4 and Q6-Q13 are more appropriate to be employed as the pairs of opposing questions

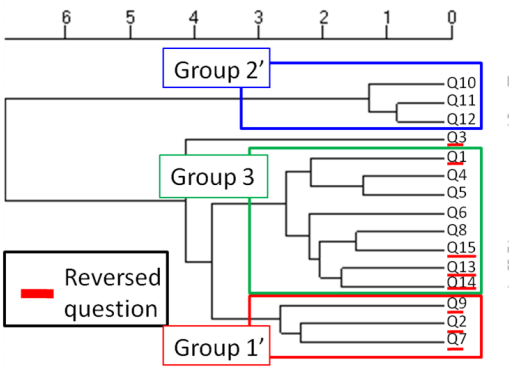


Fig. 5. Distance between questions visualized by dendrogram (after reverse)

TABLE III
RATIO OF RESPONDENTS

	Q4	Q5	Q6	Q8	Q15
Q1	0.32	0.33	0.31	0.40	0.44
Q13	0.39	0.37	0.32	0.43	0.47
Q14	0.47	0.48	0.48	0.54	0.51

(a) $C = 0$

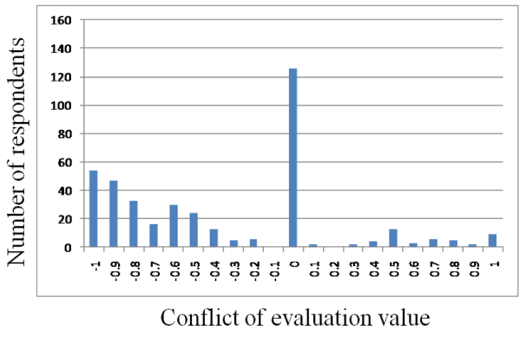
	Q4	Q5	Q6	Q8	Q15
Q1	0.12	0.18	0.21	0.21	0.16
Q13	0.09	0.12	0.11	0.14	0.14
Q14	0.12	0.15	0.17	0.11	0.12

(b) $C > 0$

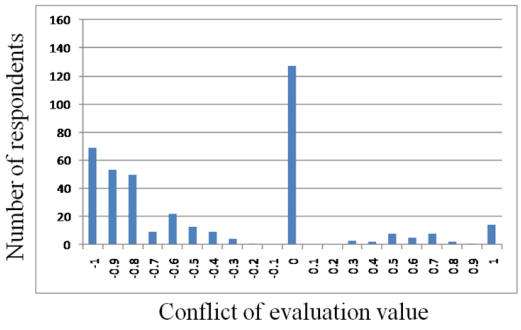
than the prepared Q4-Q13. This paper calls this derived pairs Q1-Q4 and Q6-Q13 "derived pairs." Q1 "This service will be valued by only a few people" and Q4 "I want this kind of service," and Q6 "This service has a reality" and Q13 "This service is not necessary" are difficult to be supposed as the pairs of opposing questions preliminarily though they are understandable after they were derived. It can be said that interesting result was acquired in this experiment.

D. Investigation of derived pairs of opposing questions

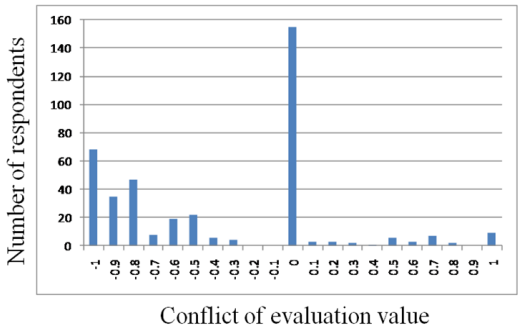
Validity test was done for the prepared pairs and the derived pairs. First, the prepared pairs (Q9-Q11, Q4-Q13) and the derived pairs (Q1-Q4, Q6-Q13) were used to calculate the conflict of response C. Here, the answers with $C > 0$ were defined as "frivolous answers" and those with $C \leq 0$ as "conscientious answers." The average of character counts in free text form were compared between these answers. Table IV shows the average character counts of frivolous answers and conscientious answers in three free-form texts. If it is assumed that frivolous respondents do not write much in the question of free-form texts, it is thought that the average of character counts in frivolous answers is smaller than that of conscientious answers when appropriate pair(s) of opposing



(a) Q1-Q4



(b) Q6-Q13



(c) Q4-Q13

Fig. 6. Histogram of conflict of response

questions are employed to calculate the conflict of response.

As shown in Table IV, when the prepared Q9-11 and Q4-Q13 were employed as the pairs of opposing questions, it does not show the difference of the average of character counts between the frivolous answers and the conscientious answers. On the other hand, when the derived Q1-Q4 and Q6-Q13 were employed, it shows obvious difference between them. The statistical t-test showed the significant difference at 5% significant level. This result shows that respondents do not necessarily regard the prepared pairs of questions as opposite

TABLE IV
AVERAGE OF CHARACTER COUNTS

	Free text 1		Free text 2		Free text 3	
	Frivolous	Conscien-tious	Frivolous	Conscien-tious	Frivolous	Conscien-tious
Prepared pairs	25.0	25.5	21.0	22.0	19.7	20.2
Derived pairs	19.0	26.4	16.9	22.6	15.5	20.8

meaning just as the questioner intended. In the proposed method, the questioner can make sure whether the respondents regarded the prepared pair(s) of questions as opposite meaning, and he/she can also extract appropriate ones from the questions without preparing.

IV. CONCLUSION

This paper defined the distance between questions [7] which showed the similarity of questions for respondents and proposed a method to derive the pair(s) of opposing questions based on them. This paper applied the proposed method to an actual questionnaire data on visualization services for next generation. It showed that the prepared pairs of opposing questions were not necessarily regarded as opposite meaning for the respondents and more appropriate pairs for the quantification of the conflict of response could be extracted, which was investigated through the comparison of the character counts in free-text form. As a future work, we will investigate the appropriate threshold to derive the pairs of opposing questions from candidates and the number of them.

REFERENCES

[1] Kuroda, S., Yoshikawa, T., and Furuhashi, T.: A Proposal for Analysis of SD Evaluation Data by Using Clustering Method Focused on Data Distribution, Proc. of the International Symposium on Frontiers of Computational Science 2005 (FCS2005), pp.317-320, 2007

[2] Maeda, S., Futatsuka, M., Yonesaki, J., and Ikeda, M.: Relationship between questionnaire surbey results of vibration complaints of wheelchair users and vibration transmissibility of manual wheelchair, Environmental Health and Preventive Medicine, Supringer Japan, Vol.8, No.3, pp.82-89, 2003

[3] Osgood, C.E., Suck, G.J., and Tannenbaum P.H.: The Measurement of Meaning, University of Illinois Press, 1957

[4] Fukami, T., Watanabe, Y, Yoshikawa, T., Furuhashi, T., Hara, I., and Yoneda H.: Discovering Minority Groups by Interactive Clustering in visible Space, Proc. of International Conference on Kansei Engineering and Emotion Research 2009, 3F(3), 2009

[5] Shu S, L.: An Internet survey for perceptions of computers and the World Wide Web: relationship, prediction, and difference, Computers in Human Behavior, Vol.18, iss.1, pp.17-35, 2002

[6] Michael D. K, Timothy D. H., and Ralph L. :A comparison of web and mail survey response rates, Public Opinion Quarterly, Vol.68, iss.1, pp.94-101, 2004

[7] Watanabe, Y, Yoshikawa., and T., Furuhashi, T.: Study on Quantification of Respondent's Conflict and Interest and Analysis Method of Questionnaire data, Journal of Japan Society of Kansei Engineering (in Japanese), Vol.9, No2, pp.129-135, 2010

[8] Watanabe, Y, Yoshikawa, T., and Furuhashi, T.: Study on Conflict of Evaluation Values and Interest of Respondent in Questionnaire Data, International Conference on Kansei Engineering and Emotion Research (KEER2010), 2C(3), 2010