A Model for Interest Measurement by Observable Non-Verbal Behavior

Sadanori Ito (sito@cc.tuat.ac.jp)[(1)(2)(3)] **Shoichiro Iwasawa[(2)(3)]** **Ichiro Umata[(2)]** **Kiyoshi Kogure[(2)]**

[(1)]Graduate School of Engineering,
Tokyo University of Agriculture and Technology
2-24-16 Nakacho, Koganeishi, Tokyo 184-8588, JAPAN

[(2)]ATR Media Information Science Laboratories,
2-2-2 KeihannaGakkentoshi,
Kyoto 619-0288, JAPAN

[(3)]ATR Intelligent Robotics and Communication Laboratories,
2-2-2 KeihannaGakkentoshi,
Kyoto 619-0288, JAPAN

Abstract

This paper proposes a model of non-verbal behavior in information acquisition activity with a focus on the differences between social and individual behavior. We derived a model from the factor analysis results of information acquisition activity for ubiquitous computing environments. Subjects were clustered by tendencies of factor scores and then compared with subjects' evaluation of the level of interest. The results reveal a correlation between non-verbal behavior and interest that depends on the statistical characteristics of the behavioral model.

Introduction

In the field of ubiquitous computing research, development is underway to create systems that can record various activities in day-to-day life or work situations as well as automatically recognizing and providing support for such activities. The goal of this research is to develop systems that do not interfere with natural human behaviors and that do not require difficult mechanical operations (Weiser, 1991). We have developed a system that infers the tendencies of people's interests in venues where people gather to acquire information or engage in intellectual exchanges, such as museums or academic conferences. Our system supports people's intellectual activities based on these tendencies. We are developing small devices that record human behaviors without input operations to infer these interest tendencies through natural non-verbal behaviors (Ito et al., 2005). Several studies have suggested methods to suppose interest and intention from non-verbal behavior (e.g., Chiu, 2000; Stiefelhagen, 2002). Our device's unique feature is that, despite its compact size, it offers a "situation recognition" function achieved by incorporating standards for detecting meaningful behavior-related information (e.g., the distance at which an object is visible to the subject) into the physical characteristics of sensors. In many cases, research in ubiquitous computing relies on devices with similar simple recognition functions (e.g., Nishimura et al., 2004). When using these devices, however, it is difficult to discover new indexes for inferring interests.

In this paper, we describe an experiment setting constructed to enable highly accurate simultaneous recording of physical actions, gazes, and utterances. Within this setting, we conducted an experiment that quantitatively evaluated non-verbal behaviors during information acquisition that reflected the subject's own interests. In this experiment, we tested the following two hypotheses. (1) When people acquire information in accordance with their own interests, their non-verbal behavior tendencies can be divided into quantitatively different groups. (2) It is more appropriate to use specific methods for inferring individual interests for each group in accordance with that group's non-verbal behavior characteristics than to use the same method for all groups. By testing these hypotheses, we will construct effective methods for inferring people's interests.

Analysis of information acquisition behavior

Experimental design

We conducted an experiment that focused on information acquisition in a space that allowed free movement. We recorded non-verbal behaviors by people involved in information acquisition activities and examined the statistically significant differences in these behavior tendencies.

Task One subject was assigned the role of a person planning a vacation and gathering information; another subject played a guide who provided information to the above subject to create a sightseeing plan that matched his own interests. To minimize differences resulting from the subjects' background knowledge and ability to understand, we selected materials related to sightseeing in locations far removed from where the subjects live.

Figure 1: Example of sightseeing panels

Setting We constructed an experiment setting to represent a sightseeing office with seven sightseeing information panels (91 x 60 cm). The panels, arranged around the perimeter of the experiment setting, provided the following information: food and drink, arts and crafts, scenic spots, art galleries, museums, historical sites, and a list of tourist information. All panels had the same configuration. Two posters were affixed to each panel, top and bottom, providing information on sightseeing spots. Each poster was then divided into two segments, right and left, providing general and detailed information (Figure 1). For example, on the "Scenic Spots" panel, the two posters provided information on "ZAO" and "MATSUSHIMA." The general information in the ZAO poster included its location and highlights and a brief introduction to the items in the detailed information segment. The detailed information described two famous spots in ZAO: "Crater Lake" and "Silver Forest."

Subject MC: Guide A professional information clerk (female, 30 years old) played the "guide" role.

Subjects S1 ~ S22: Tourists Subjects S were 12 male and 10 female students around 20 years of age who lived near the ATR laboratory.

Procedure Subject MC, who played the guide, provided sightseeing information that reflected the requests of Subject S, who played the tourist. A script was designed and information was given according to it to ensure constant quality in information. The following is an outline of the script.

1. Lead Subject S to the front of the panel, show her/him a list of sightseeing information, and ask her/him to select interesting items in order of preference.
2. To create a relaxed atmosphere, encourage Subject S to speak freely at any time.
3. In front of each guidance panel, first explain the "outline" of each poster and then provide detailed explanations in response to requests from Subject S.
4. Explain all posters, and end the presentation.

We adjusted the script so that explanations of all outlines only requires about 10 minutes; explanations of all panels including detailed explanations and questions requires about 20 minutes.

Subjects S, who played the role of tourists, were instructed to gather information from the guide at the sightseeing office and create a sightseeing plan tailored to their interests. They were told that they could move about freely and talk during the experiment. After the experiment was completed, they filled out a descriptive questionnaire in which they entered scores on a 5-point scale that asked how much they knew about the content explained in the outlines and the detailed descriptions of all posters and whether they were interested in that content. They also answered Yes/No questions regarding whether they asked questions or requested explanations from the guide and whether the answers were sufficient.

Figure 2: Experiment setting

Construction of a Measurement Environment

Figure 2 is a photo of the experiment setting. The positions of the panels and the positions and physical movements of the subjects were measured using a Vicon Motion Capture System. This system can record marker positions with a time resolution of 60 Hz and a space resolution of about 1 mm. The dimensions of the measurement space constructed for this experiment were 350 cm (w) x 450 cm (d) x 250 cm (h). Subject gaze direction was measured using Nac Image Technology Inc.'s EMR-8B, a gaze measurement device that resembles a hat. EMR-8B can record eyeball movements with a time resolution of 30 Hz and an angle resolution of about 0.15º. Subjects wore the "hat-type" gaze measurement device, a close proximity microphone, and a marker (Figure 3). To enable subjects to move freely within the experiment space, we used a flexible connecting cable 10 m long and 4 mm in diameter. All measurement devices were synchronized using the time code on the DVCAM video recorder. Synchronized information recorded included "field of view video" from the gaze measurement device, gaze point coordinates from the gaze image surface, 3D coordinates for markers obtained from a motion capture device, audio, and video.

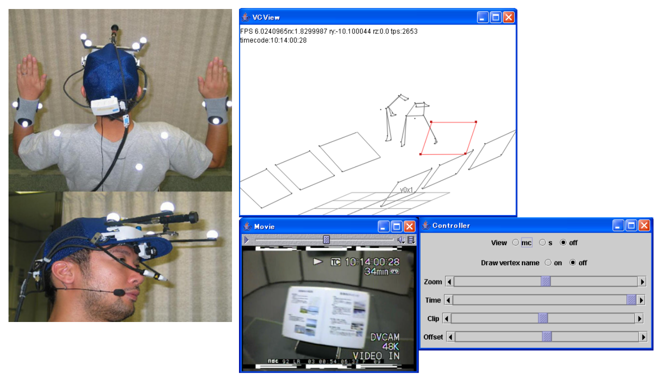

Figure 3: Measurement system and motion viewer

Markers were affixed to each subject's head (4 pts.), shoulders (3 pts.), and arms (3 pts.) and to the four corners of the panels. The four points on the head were used to determine face position and direction and the position of the gaze measurement device. The three points on the shoulders were used to determine the body's position and posture. The three points on the arms were used to detect finger pointing behavior.

The gaze point data are described in a field of view coordinate system. Motion capture marker position data are described in a global coordinate system. Using the position data from the gaze measurement device, it is possible to change the gaze point coordinates from the field of view coordinate system to a global coordinate system and derive gaze vectors in the global coordinate system. In this way, we can record, with a high degree of accuracy, what the subjects were seeing and when, as they freely moved about the experiment space.

Indexing of observation results

The non-verbal behaviors were quantified with the following indices based on observation results.

The number and the length of utterances (NU and LU) The number and total duration of utterances were recorded for each subject. Threshold values for microphone input power levels were set for each subject to segment his/her utterances.

The number and the length of gaze (NG and LG) The number and the total duration of gaze at the partner or a panel were recorded for each subject. We categorized an eye-movement sequence as an occasion of "gaze at a panel/the person" when it is within 5 degrees for more than 0.5 second.

The number of gaze movement (NGM) The number of eye-movements between each gaze was recorded while Subjects S and MC were looking at the same poster.

The length of joint attention (LJ) We recorded the length of time during which Subjects S and MC were looking at the same poster.

The distance between the gaze points (DG) The distance between the subjects' gaze points on the panel surface was measured while S and MC were looking at the same objects.

The number and the length of guided joint attentions (NGJ and LGJ) The number and total duration of the occurrences in which a gaze at a panel was overlapped or followed within 0.5 seconds by the partner's gaze were recorded.

The number of nods (NN) The number of nods was recorded for each subject. We categorized a head movement as a nod when the angle of the head changed more than 5 degrees in relation to the floor within 0.3 second.

The number of looking into the panel (NLP) The occurrences of states when the distance between the head and a panel was less than 1000 mm were counted for each subject.

The number of pointing (NP) The number of finger-pointings was recorded for each subject.

Relative distance and angle (Dist. and Ang.) Past research has indicated that the positional relationship between persons in a dialog is affected by the roles and social relationships of the persons involved (e.g., Baker, R. 1968). We therefore recorded the average values for relative distance (Dist.) and angle (Ang.) to determine the positional relationship between Subjects.

Experimental results

Three (one male and two female) of Subjects S were dismissed as evaluation subjects for missing gaze data values that resulted from an instability in gaze measurements, stemming from eyelid and eyelash conditions. The average time required by each pair for the experiments was 667.53 seconds (S.D.=158.67). The average time required for each guidance poster was 47.42 seconds (S.D.=34.74). Broken down into general and detailed explanations, 197 cases were comprised only of general explanations, with an average time of 37.50 seconds (S.D.=29.79), and 67 cases were comprised of general explanations and detailed explanations, with an average time of 76.56 seconds (S.D.=37.51). We believe that the guidance script functioned sufficiently, allowing us to obtain stable results with little difference in conditions among the guidance panels and between Subjects.

Table 1 Correlation coefficient between time length of poster guidance and the interest.

	L.P.	I.B.
D.G. (n=264)	.576**	
I.B. (n=264)	.454**	
I.D. (n=67)	.273*	.722**

Spearman Rank-Order Correlation Coefficient.
**. significance level (two-sided) P < .001
*. significance level (two-sided) P < .05

Under these conditions where the script operated smoothly, we could estimate the degree of detailed guidance information and Q&As based on the time for each poster guidance. We assumed a high correlation between the length of time and degree of interest; this correlation was confirmed by the experiment results (Table 1). Our past research that targeted exhibitions basically maintained this approach (Sumi, 2004). In general, however, it is often difficult to obtain stable results regarding length of time due to a variety of factors, including the guide's degree of skill and the content involved. For this reason, it was necessary to incorporate indexes that could be used even when non-verbal behavior detection results were normalized by length of time.

Detection results for non-verbal behavior Table 2 shows the detection results for non-verbal behavior derived from each pair in the experiment. In almost all indexes detected, there are clearly broad individual differences in the behaviors of student Subjects S compared to the guide MC.

Clustering of non-verbal behavior tendencies Non-verbal behavior indexes for Subjects S1 through S22 were normalized using guidance times for each poster (P1-P12). Table 2 shows the correlation coefficients between these results and the degree of interest. We identified an intermediate level of correlation between the social behavior of utterances and attention on the MC, but it was insignificant.

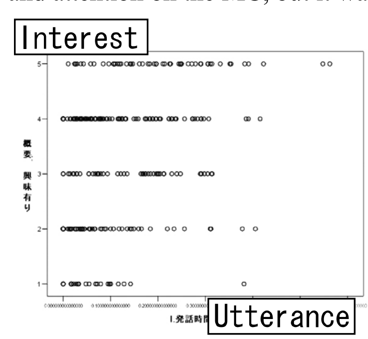

Figure 4: Scatter chart by utterance time length and interest degree

Table 2 Detected non-verbal behavior and correlation coefficients

		L.P.	N.U. s	N.U. Mc	L.U. s	L.U. mc	N.G. s-mc	N.G. s-p	N.G. mc-s	N.G. mc-p	L.G. s-mc	L.G. s-p	L.G. mc-s	L.G. mc-p	Dist. s-mc	Ang. s-mc
Mean			96.05	49.86	107.75	597.61	49.32	137.45	108.14	214.09	74.91	522.63	139.77	489.26	957.88	171.33
S.D.			52.94	13.94	71.22	145.76	44.43	62.65	47.73	77.37	131.32	114.86	64.25	139.72	116.76	9.89
C. C.	I.B.	.454**	.398**		.404**		.381**	.233**			.373**	.258**			-0.120	.149*
	I.B. Nrm.		.275**		.334**		.337**	0.004			.364**	-.127*				

		L.J. s-mc	D.G. s-mc	N.G.J. s-mc	N.G.J. mc-s	N.N. s	N.N. mc	N.P. s	N.P. mc	N.B. s	N.B. mc	N.L.P. s	N.L.P. mc	N.G.M. s	N.G.M. mc	
Mean		14.16	312.06	67.14	116.36	81.77	325.05	22.68	51.14	211.32	266.00	6.55	10.50	76.27	80.09	
S.D.		4.35	101.72	40.28	37.12	61.58	80.52	77.96	27.64	145.11	69.27	19.57	6.97	40.68	24.84	
C. C.	I.B.	.268**	.153*	0.124	.256**	.308**		0.019		.157*		-0.096		.166*		
	I.B.Nrm.	-.277**		-0.068	-.167**	.215**		.155*		-0.077		.131*		-0.014		

Spearman Rank-Order Correlation Coefficient (time scale: sec., distance scale: mm)
** $p < 0.01$, * $p < 0.05$
(I.B.: interest in brief guidance, I.D.: interest in detailed guidance, L.P.: length of time guided with a poster, N.U.sub.: number of utterances, L.U.: length of utterances, N.G.a-b: number of gazes, L.G.: length of gazes, L.J.s-mc: length of joint attentions, D.G.s-mc: distance between gaze points, N.G.J.: number of guided joint attentions, N.N.: number of nods, N.P.: number of pointings, N.B.: number of blinkings, N.L.P.: number of times looking at a poster, N.G.M.: number of gaze movements)

For example, as seen from the scatter chart of utterance time length and degree of interest (Figure 4), a longer utterance time suggests a higher degree of interest, but we cannot say the reverse; in other words, this correlation is not strong enough to estimate degree of interest from non-verbal behavior. We also cannot determine degree of interest from individual behaviors such as focus of attention on posters.

We therefore conducted a factor analysis using Varimax rotation principal component analysis to determine the common factors in Subject S's non-verbal behavior. We extracted four factors with a primary eigenvalue of 1 or more. The cumulative contribution rate for these four factors was 57.07%. Table 3 shows the list of factors following Varimax rotation.

Table 3 Factor analysis results of non-verbal behavior

	I	II	III	IV	Communalities
N.U.s	0.020	0.972	0.214	-0.012	0.992
L.U.s	0.057	0.812	0.172	-0.010	0.692
N.G.s-mc	0.418	0.071	0.743	-0.079	0.739
N.G.s-p	-0.111	0.237	0.390	0.630	0.617
L.G.s-mc	0.882	-0.114	0.345	-0.116	0.923
L.G.s-p	-0.268	-0.007	-0.037	0.465	0.289
L.J.s-mc	-0.888	-0.183	-0.174	0.028	0.854
N.N.s	0.058	0.263	0.685	-0.147	0.563
N.P.s	0.740	-0.010	-0.218	0.014	0.596
N.B.s	0.180	-0.010	0.272	0.241	0.165
N.L.P.s	-0.063	0.056	0.086	-0.205	0.056
N.G.M.s	-0.157	0.236	0.521	0.107	0.363
Factor contribution	2.442	1.840	1.800	0.766	
Cumulative contribution	20.352	35.689	50.691	57.074	

Extraction Method: Principal Factor Method
Rotation method: Varimax with Kaiser Normalization

One unique feature of Factor I is the high factor load volume for length of eye contact time when looking at the MC and for the number of finger pointing actions. Another is behaviors with social functions, such as conveying understanding of the contents based on the MC's guidance, or Subject S conveying topics or intent to the MC. On the other hand, the length of joint attention (when both MC and S were looking at the same poster) was high in the negative direction, indicating an absence of relaxed behaviors by the listener in following the text or images in the posters in reflecting the MC's verbal guidance. A unique feature of Factor II is the high factor load for number and length of utterances, indicating social behaviors aimed at gathering information or promoting smooth dialog through supporting responses, questions, and other utterances. One unique feature of Factor III is the high factor load volume for factors with weak response-related social functions (i.e., responses less active than utterances) directed at the MC; for example, the number of eye contacts focusing on the MC or the number of nods. There was also a high factor load for the number of gaze shifts and a low factor load for joint attention, which suggests a combination of social behaviors and independent behaviors, such as looking at posters independent of the MC even while continuing to make weak responses to the MC. A unique feature of Factor IV is the high factor load for number and length of gazes at the posters. This suggests a tendency toward information acquisition activities with an emphasis on individual behaviors.

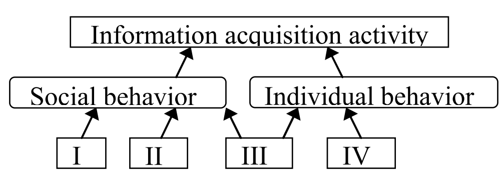

Figure 5: Model of information acquisition activity

In this way, we extracted both social and individual behaviors as factors from the measurement results of behavior indexes for non-verbal behaviors. In Figure 5, we show our proposed model for non-verbal behaviors in information acquisition activities based on these results.

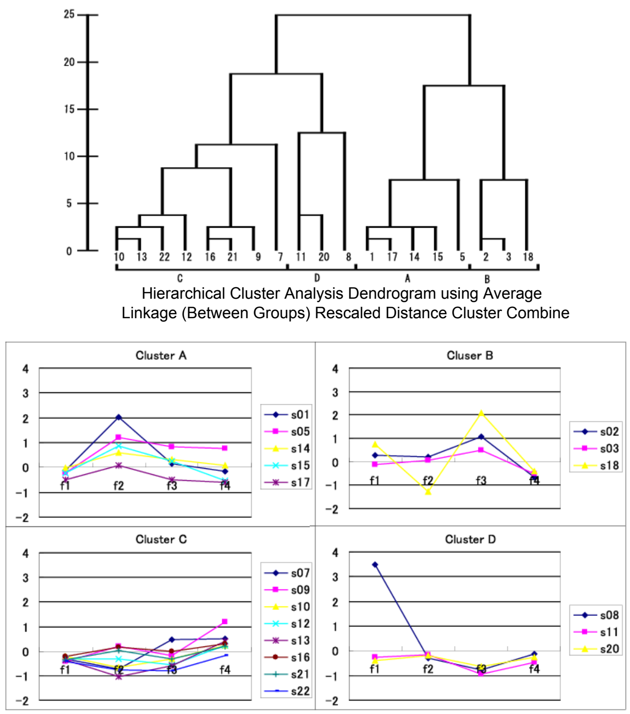

Figure 6: Dendrogram and factor scores of cluster analysis results

We conducted a hierarchical cluster analysis to tentatively categorize subjects based on non-verbal behaviors during these information acquisition activities. Figure 6 shows a dendrogram obtained by calculating average factor scores for each subject and applying Pearson correlations and an average linkage between groups. In this figure, we used four clusters to clearly indicate the characteristics of Subject groups and showed factor score distributions for Subject S's in each cluster. The main feature of Cluster A is the high factor scores for social behaviors, particularly Factor II related to utterance behaviors. The main features of Cluster B are high scores for Factor III, social behaviors related to response elements, and low scores for Factor IV, the number of gazes at posters (an individual behavior). A notable feature of Cluster C was the concentration of each factor score around the 0 mark and a minimal amount of variation between factors. In Cluster D, S08 showed an extremely high score for Factor I, but if we consider this an exception, a feature for the rest of this Cluster was a low score for Factor III, which indicates response-related social behaviors such as number of gazes and nodding. To summarize, Cluster A had high factor scores for active social behaviors, cluster B had high factor scores for responses and other pseudo-active social behaviors, cluster C had low scores overall with very little variation in factor scores, and Cluster D had low factor scores for response-related behavior.

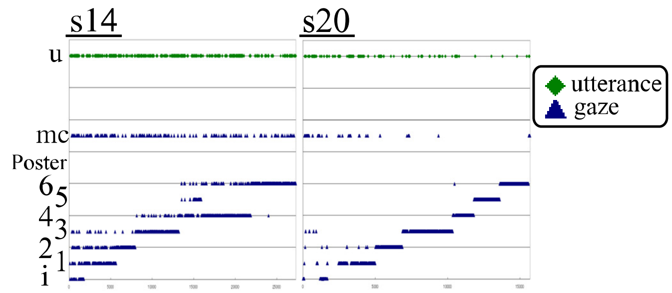

Figure 7: Examples of non-verbal behavior of subjects in clusters A and D

Figure 7 shows two examples of the plots of non-verbal behavior during the experiment: Subject 14, from Cluster A, and Subject 20, from Cluster D. The average degree of Subject 14's interest in all posters is 3.43 (on a 5-point scale) and Subject 20's average is 3.30, so there is no significant difference between them. From this plot, we can see that Subject 14 demonstrated extremely social and active behaviors, for example, many utterances and frequent gazes at the MC. Subject 20 demonstrated few utterances and vastly fewer glances at the MC compared to Subject 14. Note, however, that attention on the poster was very stable and reflected the order of the MC's guidance. We can thus assume that other indexes are required to estimate the degree of these two Subjects' interest from their non-verbal behavior.

Table 4 Correlation between non-verbal behavior and interest in brief guidance

Cluster	N.U. s	L.U. s	N.G. s-mc	N.G. s-p	L.G. s-mc	L.G. s-p	L.J. s-mc	N.N. s	N.P. s	N.G.M. s
A	.243	.476**	.529**	.074	.572**	-.211	-.453**	.051	.151	-.135
B	.187	.269	-.192	.147	.008	.369*	.093	.145	.313	-.481**
C	.199	.285**	.299**	-.021	.355**	-.155	-.333**	.353**	.040	-.088
D	.306	.140	.064	.104	.036	-.030	.120	.044	-.209	.485**

Cluster: A (n=60), B (n=36), C (n=96), D (n=36)
Spearman Rank-Order Correlation Coefficient
** $p < 0.01$, * $p < 0.05$

We recalculated the correlation coefficients of non-verbal behavior and the degree of interest using clustering results (Table 4). We confirmed that Cluster A, which had high social factor scores, also demonstrated high correlation coefficients between degree of interest and social behaviors such as length of utterances and length of attention on the MC. We believe that Subjects who emphasize such social functions in the dialog process have a high tendency to express their own interests to others. Similar results can also be seen in Cluster C, which showed low variation in factor scores, although the correlation coefficient was not high. In Clusters B and D, we can see a correlation with attention on the posters and gaze shift, which are individual behaviors. In Cluster B, we observe low correlation with attention on the poster, and simultaneously, a negative correlation with gaze shift. Perhaps the Subject was focusing attention on an item distinct from the poster being explained. As we saw in Table

1, the correlation between these individual behaviors and interest tendencies could not be confirmed using existing methods and appear as a result of a clustering effect.

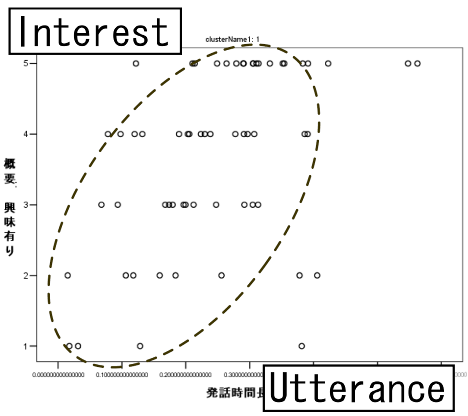

Figure 8: Scatter chart by utterance time length and interest degree of cluster A

Figure 8 shows a specific example of clustering in a scatter chart of utterance time length and degree of interest for Cluster A. Here, we can see that compared to Fig. 5 (before clustering), Subjects with few utterances regardless of a high degree of interest have been removed, which increases the reliability of the estimates of interest.

Conclusion

In this paper, we constructed an experiment setting that enabled accurate synchronization and recording of gaze, physical behaviors, and utterances as non-verbal behaviors during information acquisition activities. Using this setting, we conducted an experiment to quantitatively evaluate non-verbal behaviors during information acquisition that reflected the Subject's own interests. We conducted a factor analysis of quantitative results for non-verbal behaviors derived from the experimental results and proposed a model of non-verbal behaviors in information acquisition activities comprised of four factors representing social and individual behaviors. We then clustered S Subjects in accordance with this model, based on tendencies in non-verbal behaviors. After summarizing the correlations between degrees of interest and non-verbal behaviors, we derived a higher correlation coefficient than without clustering. We were also able to observe a correlation between degree of interest and individual behaviors such as gazing at posters, which could not be obtained with other methods that do not use clustering. In terms of non-verbal behaviors that express interest, it became clear that there is a difference between groups that show many individual behaviors and groups that show many social behaviors in the process of information acquisition activities.

After establishing more subtle quantitative indexes for these non-verbal behaviors in the future, we will create indexes for social protocols with a focus on the temporal continuity of mutual behaviors and conduct analyses based on these indexes. We will also create settings for observing non-verbal behavior using more simplified detection devices and incorporate these settings into practical and applicable models.

Acknowledgments

This research was conducted as part of "A study of innovational interaction media toward a coming high functioned network society," with funding from the National Institute of Information and Communications Technology (NICT).

References

Baker, R. (1968). *Ecological Psychology: Concepts and Methods for Studying Human Behavior*. Stanford University Press, Stanford.

Bono, M., Suzuki. N., Katagiri Y. (2004) An Analysis of Participation Structures in Multi-Party Conversations: Do interaction behaviors give clues to know your interest? *Cognitive Science*, No.11, Vol. 3, 214-227. (in Japanese)

Chiu, P., Kapuskar, A., Reitmeier, S., Wilcox L. (2000). Room with a rear view: Meeting capture in a multimedia conference room. *IEEE Multimedia Magazine*, Vol. 7, No. 4, pp. 48-54.

Ito, S., Iwasawa, S., Tsuchikawa, M., Sumi, Y., Mase, K., Katagiri, Y., Kogure, K., Hagita N. (2005). Implementation and Evaluation of Wearable Sensor Devices for Recording and Detecting Conversational Interaction. *in Proc. of ISWC2005*, pp. 130-137.

Nishimura, T., Nakamura, Y., Itoh, H., Nakashima, H. (2004) System Design of Event Space Information Support Utilizing CoBITs. *in Proc. of ICDCS Workshops 2004*, pp. 384-387.

Stiefelhagen, R., Yang, J., Waibel, A. (2002). Modeling focus of attention for meeting indexing based on multiple cues. *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 928-938.

Sumi, Y., Ito S., Matsuguchi, T., Fels, S. Mase, K. (2004). Collaborative capturing and interpretation of interactions, *in Proc. of Pervasive 2004 Workshop on Memory and Sharing of Experiences*, pp.1-7.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, Vol. 265, No. 30, pp. 94-104.