Toward Interest Estimation from Head Motion Using Wearable Sensors: A Case Study in Story Time for Children

Ayumi Ohnishi[1], Kaoru Saito[2], Tsutomu Terada[1,3(✉)], and Masahiko Tsukamoto[1]

[1] Graduate School of Engineering, Kobe University, Kobe, Japan
`tsutomu@eedept.kobe-u.ac.jp`
[2] Graduate School of Frontier Science, The University of Tokyo, Tokyo, Japan
[3] PRESTO, Japan Science and Technology Agency, Kawaguchi, Japan

Abstract. Learning activities were evaluated using questionnaire survey, video, or audio. However, these methods have the following problems. First, writing on the questionnaire paper is difficult, especially for little children. Second, because the answering questionnaires was performed after the experiments were finished. They were different temporally and spatially from the scene to be evaluated. Moreover, sometimes participants has forgot the part of the contents. Finally, by recording video or audio, we can look back at each scene and evaluate them, however, video or audio analysis takes a very long time. This research aims to solve these three problems and evaluate natural reactions; first, for children of a low age group, second, including changes in the state of participants during an activity, and third, as much as possible without wasting time and effort. In this paper, during storytelling events for children, we attempted to obtain the values of acceleration and angular velocity sensors with sensors placed on the participant's heads, and tried to estimate their motions and degree of interests. Motions were calculated using the F-value, with accuracies of 0.66 in "Sitting state", 0.26 in "Sitting again", 0.47 in "Wriggling", and 0.93 in "Playing with hands". From these results, "Playing with hands" had the highest degree of interest, with a motion recognition rate of 0.93 in F-value. Comparing the proposed method with the video evaluation later, the proposed method can obtain the evaluation result during the learning activity. Therefore, by feeding back the estimation result in real time, we can make improvements while doing activities.

Keywords: Motion recognition · Interest estimation · Book reading to children · Wearable sensing

1 Introduction

After conducting learning activities such as lessons and little learning, by evaluating these activities and designing based on the evaluation results, the next activities are further improved. To evaluate methods of these activities are generally used questionnaire surveys, video, and audio recordings. However, as shown in Table 1, there are three problems described as follows.

Table 1. Score sheet of each evaluation method in learning activities

	(a) Evaluation of children	(b) Evaluation dividing scenes	(c) Time and effort to evaluate
Questionnaire	Difficult	Difficult	Quickly, and easily
Video, or audio	Evaluable	Evaluable	Taking a long time
Wearable sensing	Evaluable	Evaluable	Quickly, and automatically

(a) Questionnaire surveys can analyze trends against many participants at a time, however, it is difficult for little children to answer.

(b) The participants are evaluated after the objective is done. Therefore, it is different temporally and spatially from the scene to be evaluated. As a typical questionnaire method, people participate in some event and answer the questionnaire after all event contents are over. Therefore, in this method, people may be more impressed with the contents of the last one than the first one. As a result, some questionnaire answers are influenced by the contents of the last one.

(c) In participant observation, an analysis after the examination is a waste of time. Furthermore, there is a possibility of giving an oppressive feeling to participants by increasing the number of cameras used to record personal detailed movements from multiple angles. It is common to receive data related to physical appearances, facial expressions, and gestures from images obtained by video cameras.

This research aims to solve these problems and evaluate natural reactions; (a) for children of a low age group, (b) including changes in the state of participants during an activity, and (c) as much as possible without wasting time and effort.

Then, consider which method is actually suitable as an evaluation method of activities. Yamashita et al. attempted to evaluate activities by implementing Sounding Board. This system records a person's assessments in real activities such as conversations [1]. However, it is difficult to use it on a daily basis, because it is necessary to point the PDA terminal to the participant to be assessed, and to operate it. By accumulating everyday casual person's gestures, such as nodding or neck cranking, we can analyze the data from that content. However, these evaluations are improvised and not recorded. Therefore, if we record this casual evaluation, it possibly can be used as an indicator of interest during learning in addition to the conventional evaluation such as a questionnaire. Therefore, we propose a method to analyze evaluation activities by wearable sensing. As shown in Table 1, wearable sensing can solve (a) to (c) problems. In details, (a) it becomes possible to evaluate from the reaction of children, (b) time series analysis is possible, and (c) applications using this analysis method can evaluate quickly and automatically. In this paper, during storytelling events for children, we acquired acceleration and angular velocity from the subjects that participated with a cap with a motion sensor attached. At the same time, one movie recorded for annotation. Using the acceleration and angular velocity data, we attempted to recognize their reactions and estimate degree of interests from these natural motions. We evaluated by the following steps. First, we calculated the recognition accuracy of the motion seen in

the story time. Second, we created five degrees of interest indicators. Additionally, two observers judged the degree of interest of the children to the story time from the recorded movie. Finally, the correspondence between observers' judgements and index of interest were compared, and the expressivity of the index was considered. By calculating the interest estimation rate by such a procedure, we evaluated whether each motion is effective as index of the degree of interest.

This paper is organized as follows. Section 2 describes related works. Section 3 explains the system requirements. Section 4 describes the evaluation. Finally, Sect. 5 provides the conclusion and mentions future work.

2 Related Works

Various systems that analyzes people's behaviors in conversation and recognition of head movements have been proposed. As an analysis of multiparty interaction, the sociometer implemented by Choudhury et al. is a portable device consisting of a microphone, acceleration sensor, infrared module, and GPS. They aimed to visualize the social relationships of multiple people from data obtained from gestures and conversation [2]. The Augmented Multi-party Interaction (AMI) project aims to develop meeting browsing and analysis systems. His meeting corpus is recorded by using a wide range of devices including close-talking and far-field microphones, individual and room-view video cameras, a projector, a whiteboard, and individual pens, and so on [3]. SUMI et al. developed IMADE environments to collect various kinds of information during a conversation, such as a subject's motion, gaze, voice, and biological data [4]. Tung et al. implemented a multimodal system, which consists of a large display attached to multiple sensing devices, to obtain individual speech and gazing directions [5]. Mana et al. proposed a multimodal corpus system with automatic annotation of multi-party meetings using multiple cameras and microphones. They investigated the possibility of using audio-visual cues to automatically analyze social behaviors and to create a system to predict personality characteristics [6]. Okada et al. attempted to classify nonverbal patterns with gestures, e.g., utterance, head gesture, and head direction of each participant, by using motion sensors and microphones. In this research, we are targeting events that already exist, therefore in order to take as much natural behaviors as possible, it is necessary for it to be a location- independent system. We determined that more natural evaluation behavior would be acquired by using wearable sensors. As mentioned above, there was much research conducted for estimating interests and the degree of concentrations from estimations using the camera images and motion sensors. However, a few studies evaluated behaviors and interests of young or low-grade school children. It is difficult to evaluate quantitatively for children by using a questionnaire because of the difficulty filling them. Therefore, in this research, we aim to use a story time activity targeted to such children of such an age and determine what evaluation behaviors should be measured with wearable sensors.

3   System Requirements

The assumed activity is a storytelling of picture books, which is one of the learning activities for young children, for example, like a story time. An experiment at a story time event for children held monthly was conducted at The Mount Fuji Research Institute, Yamanashi Prefectural Government. During in this story time, individual children determined their seating positions and some infants sat on their part's lap. Obtaining images from the front of each person's face with the video camera would be difficult.

To evaluate the natural behavior of children, it was necessary to bring them closer to their usual activities. As a result, wearable sensors, independent on the location were adopted. In this research, children from various elementary school lower grades, who could not answer the questionnaire, were targeted. An experiment was conducted during the story time part of the event aimed at enhancing interest in nature, with 'interest' in this activity being the main focus. Actions indicating aggressiveness and passivity from the children's behavior were detected, assuming them as an evaluation behavior, and whether it is an index of interest or not were examined. To clarify actions that could be indicators for the interest from actions the following experiments were conducted.

4   Evaluation

To estimate the degree of interest for contents, this study aims to examine behaviors which can be an index of interest in the activity during story telling. In other words, actions associated with the degree of interest and actions that can be detected with a high level of accuracy using a wearable sensor were considered.

4.1   Procedure

Evaluation experiments were conducted in the story time period cooperated with this research, which contained three different stories and time to play with hands. This event was approximately 20 to 30 min in total. Participant's children wore a cap with ATR's TSND121 sensor [8] as shown in Fig. 1 and attached acceleration and an angular velocity sensors to the right side of the cap. Moreover, as shown in Fig. 2, the children sit facing the storyteller, the staff member who reads picture card's and shows large picture books at the front. Five caps with sensors were prepared and motions of up to five heads of participants were acquired. The measurement frequency of the sensor was 20 mm/s. Table 2 lists the subjects that participated in the experiment. Test subjects included 14 children from kindergarten or lower grades of elementary school. 11 out of 14 children wore the cap until the last story. A movie for confirmation was taken by a video camera. The data was annotated from the video using Elan software [9], and correct answer data were collected. The correlation between motion recognition results based on the acceleration value/angular velocity value and correct answers was compared and evaluated.
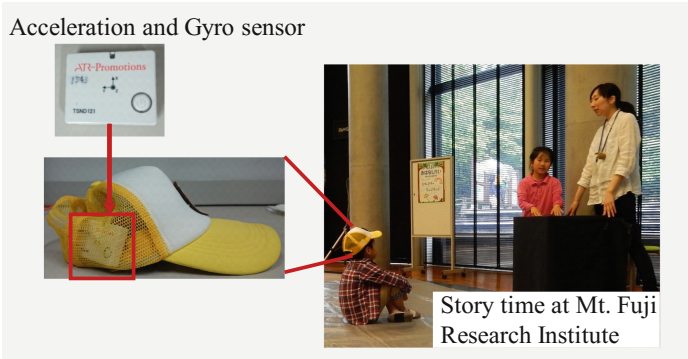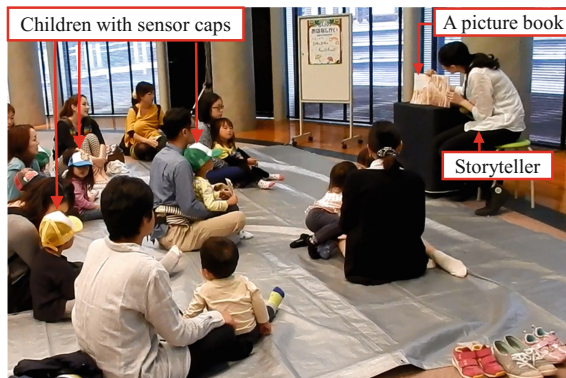
Fig. 1. Sensor position

Fig. 2. Snapshot of experiment environment

Table 2. Breakdown of the subjects in experimentation

	June 21, 2015		July 12, 2015	
	1st time	2nd time	1st time	2nd time
All subjects [persons]	5	1	5	3
Valid subjects [persons]	2	1	5	3

*Valid subjects mean full data acquired. Some children took off their caps with the sensor during the story time.

4.2 Result

Motion Recognition

Table 3 shows the motion recognition rate by acceleration/angular velocity, against a list of all actions seen during the story time. Evaluation items were six-dimensional feature values, (1) three-axis acceleration, (2) three-axis angular velocity, (3) three-axis

Table 3. Motion recognition rates

Child's actions	Recall	Precision	F-value
Sitting state	0.54	0.84	0.66
Sitting again	0.61	0.17	0.26
Wriggling	0.63	0.38	0.47
Playing with hands	0.96	0.90	0.93
Looking around	0.33	–	–
Clapping	–	–	–
Nodding	–	–	–
Finger pointing	–	–	–
Looking down	–	–	–

acceleration composite, (4) average of the composite value over one second, (5) variance of the average values, and (6) inclination angle of three-axes. The recognition rate was evaluated by 10 using division cross with validation with Weka's J48 algorithm [10]. Motions were calculated using the F-value, with accuracies of 0.66 in "Sitting state", 0.26 in "Sitting again", 0.47 in "Wriggling", and 0.93 in "Playing with hands". It was not possible to acquire gestures with occurrence frequencies less than 1% of the total time. To resolve this issue, an algorithm, which could recognize more accurately the motions seen in this experiment was needed.

Interest Judgement from Observers

Table 4 summarized the five interest degree evaluation criteria and the actual observed actions. The degree of interest from the children was evaluated continuously in five levels by two observers. Figure 3 shows the distribution of the observers' evaluations. The interest level 3, which was positioned between the high and low levels, occurred for approximately 20% the experiment duration. Figure 4 compares the evaluation of the interest degree of two observers in time series. In this case, the interest level could be used as an interest indicator; because the observer's judgements were almost always level 5 during the period of playing with hands. Degrees of interest are subjective and it is necessary to ascertain whether they are consistent with each other. Therefore, Cohen's secondary weighted kappa statistics [11] were used to confirm the reproducibility between the two evaluators. The reproducibility between the two evaluators was 0.93. It could be said the rate was almost matched in this case.

Table 4. Stage of five interest levels and participant's behavior observations

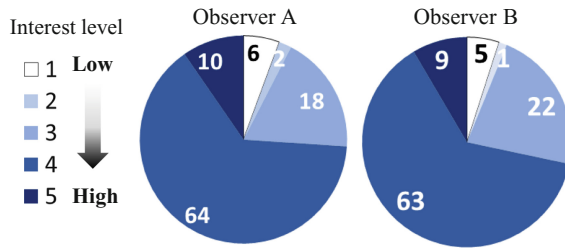|	5	4	3	2	1
Interest level	Obviously interested	Listening, seems interesting	Undecided	Does not seem to be listening	Obviously not interested
Observed actions	Joining, answering the questions, nodding	Listening, watching the speaker, clapping	Sitting state, undecided	Looking down, sitting again, wriggling	Looking around

Fig. 3. The distribution of observers' evaluation
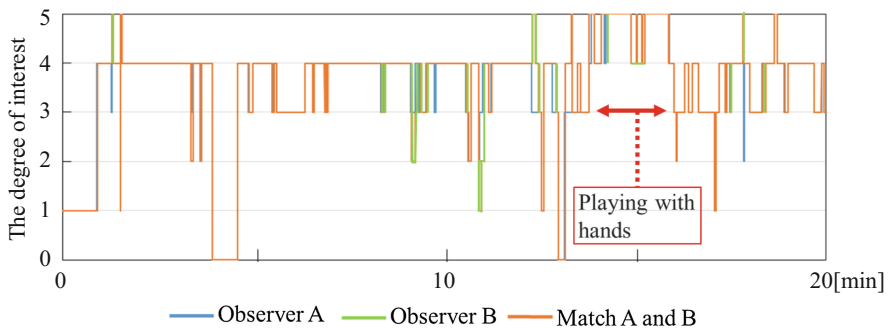
*The break time was excluded from the analysis and set to 0 degree.

Fig. 4. Interest degree evaluated from observers

Interest Estimation by Motion Recognition

The correspondence between subject's behaviors and index of interest were compared, and the expressivity of the index in Table 3. was considered. In other words, the judgment accuracy of interests in this section was evaluated. Table 5 shows the calculated results of the interest estimation accuracy from motions. Moreover, Table 6 lists the number of interests with which the index and observer's judgment matched.

Table 5. Accuracy of interest determination from motions

Child's actions	The degree of interest	Recall	Precision	F-value
Sitting state	4	0.96	0.86	0.91
Sitting again	2	0	0	0
Wriggling	2	0	0	0
Playing with hands	5	1	0.36	0.53
Looking around	1	0.38	1	0.55
Clapping	4	0	0	0
Nodding	5	0	0	0
Finger pointing	5	–	–	–
Looking down	2	–	–	–

Table 6. Comparison of index and observer's judgment

Child's actions	The degree of interest					
	Index in Table 3	Observers' judgment [times]				
		5	4	3	2	1
Sitting state	4	1	25	–	–	–
Sitting again	2	3	1	8	–	–
Wriggling	2	2	2	29	–	–
Playing with hands	5	–	–	–	5	3
Looking around	1	–	–	–	–	–
Clapping	4	–	–	4	–	–
Nodding	5	–	1	–	–	–
Finger pointing	5	1	–	–	–	–
Looking down	2	–	–	–	–	1

"Sitting state" means sitting facing forward and hardly moving. The cell of the index in Table 6 is the value based on 5 degrees of interest in Table 4. Comparing the rating results of the indices in Table 5 with the results of the observers' judgement, we found that the observer judged 'listening (4 degrees of interest of 5 stages)' when children were sitting still and looking for storyteller. When actions appeared, the things that were all rated high in interest were; "Sitting state", "Play with hands", "Nodding", and "Pointing". By contrast, the behaviors that were all rated low were "Looking around" and "Looking down". In addition, comparing index and judgment, the results for the number of "Sitting again" and "Wriggling" were scattered. For both of these motions, any positive judgment occurred before the playing with hands time. As a result, it is thought that sitting again, like the preliminary action of the play, is rated as interesting. Because it is thought that this can be seen in the children who are motivated to play, the observer judged it as interested. As for the other "Sitting still" and "Wriggling", it appeared collectively in the degree of interest 3, therefore, it could be used as an indicator of the degree of interest 3. From the results of this experiment, comparing the interest degree estimation rates of Table 5 with the motion recognition rates of Table 2, children's action, "Sitting state" and "Playing with hands" were considered to be effective as an indicator of interest. To summarize the results, the indices shown in Table 7 were obtained from the results obtained in this paper.

The correspondence between the actual actions and the indices of interest was organized, and we investigated whether it could be an indicator of interest. The

Table 7. New index based on evaluation results

	5	4	3	2	1
Interest level	Obviously interested	Listening, seems interesting	Undecided	Does not seem to be listening	Obviously not interested
Observed actions	Joining (ex. playing with hands), pointing	Sitting state, nodding	Wriggling, sitting again, clapping	Looking down	Looking around

situation in "Sitting state" and in playing with hands were effective as action indices. The hand-playing action seemed to be adopted intentionally in this event to prevent children getting bored. However, it was considered that recognition of actions in the participatory situation, besides the hand-playing action, using the wearable sensors was effective as a method of evaluating the degree of interest.

4.3 Discussion

The recognition accuracy of motion in Table 2; F-value 0.66, for "Sitting state" was lower than expected. It was influenced by the low in difference between "Sitting state" and "Wriggling". The situation observers saw in the video and judged as "Wriggling" was primarily when participants were swinging or moving their fingers. The low recognition accuracy value was caused by the position of the sensors on the participants, which did not reflected the information from their hands or fingers. In this experiment, it was not possible to use all of the motions for judging the degree of interest in the evaluation of the motion and the interest; because there were cases in which there were combined reactions such as answering while wriggling. Analysis of the state "Wriggling" was needed in detail. The state of the subject observed from the video is described as follows. First, co-occurrence relationships between the voice and the motion were seen in each subject, in many cases, some actions against sounds were recorded in the scene observers selected. Second, from the time series showing the variance of the acceleration and the degree of interest, shown in Fig. 5, changes in behavior were seen when the rating of the degree of interest changed. Finally, during this story time, the number of occurrences of actions considered to have an interest, such as agreement, was low. In regard to storytelling, improvement of the recognition rate of particular motions as indicators and judging those who clearly behave differently from the other children by
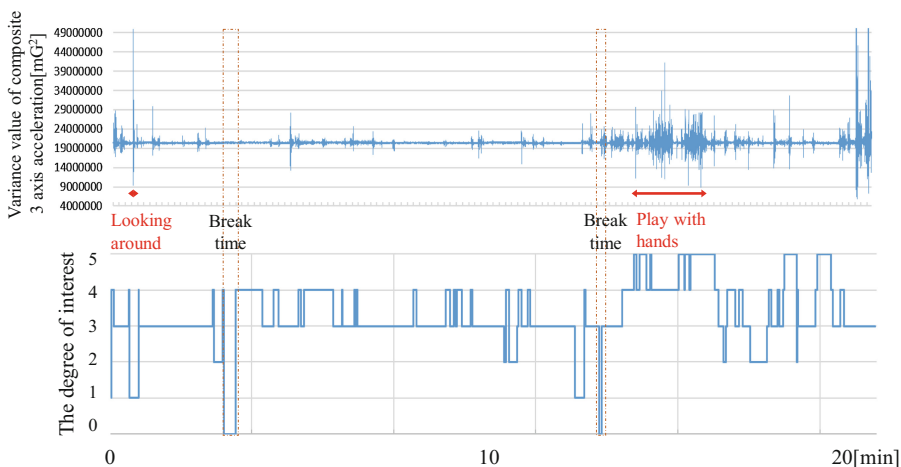
*The break time was excluded from the analysis and set to 0 degree.

Fig. 5. Comparison of acceleration value and degree of interest

using the degree of movement as an index are needed. The degree of the interest was not directly judged by the sensor values. To determine whether children were really interested in the content, further experiments were needed, for example, measuring the reaction when people showed the children obviously funny or stupid content. However, we suggested that motion recognition from acceleration and angular velocity could be used for interest estimation for the playing with hands. This research performed experiments at storytelling events for children and contributed to evaluating the degree of interests for each motions and the recognition rates of each.

5   Conclusion

In this paper, we proposed a method to acquire head movement measurements using wearable sensors and estimate interest based on its acceleration and angular velocity. As a case study, we evaluated the recognition accuracy of storytelling events targeted for children. We obtained video for confirmation, and used it to create a correct answer label of motions and degrees of interest. We then attempted to estimate the behavior and degree of interest of participants from the features of acceleration and angular velocity. Correct answer data of interests were evaluated by two observers in five levels from the video. From this experiment, ratings of the two observers' evaluation data matched by Cohen's Kappa coefficient.

As a result of the evaluation, nine motions were observed from the video data as follows; "Sitting state", "Sitting again", "Wriggling", "Playing with hands", "Looking around", "Clapping", "Nodding", "Finger pointing", and "Looking down". Moreover, four out of nine motions were recognized by using wearable sensor values, whose F-values were accuracies of 0.66 in "Sitting state", 0.26 in "Sitting again", 0.47 in "Wriggling", and 0.93 in "Playing with hands". In this case, "Playing with hands" was the highest degree of interest, with a motion recognition rate of 0.93 in F-value. To estimate the degree of interest, this paper has not reached a direct judgment from acceleration and angular velocity values. This research experimented at a storytelling event for children and contributed to the evaluating the degree of interests for each motion and the recognition rates. In the future, we will attempt to adopt an algorithm which can recognize the motions seen in this experiment more accurately. Getting a higher recognition rate by using wearable sensors enables the use of non-restricted place to estimate interest more casually.

References

1. Yamashita, J., Kato, H., Ichimaru, T., Suzuki, H.: Sounding board: a handheld device for mutual assessment in education. In: Extended Abstracts on Human Factors in Computing Systems (CHI 2007), pp. 2783–2788 (2007)

2. Choudhury, T., Pentland, A.: Sensing and modeling human networks using the sociometer. In: Proceedings of 7th IEEE International Symposium on Wearable Computers (ISWC 2003), p. 216 (2003)

3. Carletta, J., et al.: The AMI meeting corpus: a pre-announcement. In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 28–39. Springer, Heidelberg (2006). doi:10.1007/11677482_3

4. Sumi, Y., Yano, M., Nishida, T.: Analysis environment of conversational structure with nonverbal multimodal data. In: Proceedings of International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010), p. 44 (2010)

5. Tung, T., Gomez, R., Kawahara, T., Matsuyama, T.: Multiparty interaction understanding using smart multimodal digital signage. IEEE Trans. Hum.-Mach. Syst. 44, 625–637 (2014)

6. Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., Zancanaro, M.: Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In: Proceedings of International Conference on Multimodal Interfaces and the 2007 Workshop on Tagging, Mining and Retrieval of Human Related Activity Information (ICMI-TMR 2007), pp. 9–14 (2007)

7. Okada, S., Bono, M., Takanashi, K., Sumi, Y., Nitta, K.: Context-based conversational hand gesture classification in narrative interaction. In: Proceedings of 15th ACM on International Conference on Multimodal Interaction (ICMI 2013), pp. 303–310 (2013)

8. ATR-promotions. http://www.atr-p.com/products/TSND121.html

9. ELAN. https://tla.mpi.nl/tools/tla-tools/elan/

10. Weka 3, The University of Waikato. http://www.cs.waikato.ac.nz/ml/weka/

11. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measur. 20(1), 37–46 (1960)