



# Body Movements for Communication in Group Work Classified by Deep Learning

Hiroaki Sakon and Tomohito Yamamoto<sup>(✉)</sup>

Department of Information and Computer Science, College of Engineering,  
Kanazawa Institute of Technology, 7-1 Oogigaoka,  
Nonoichi, Ishikawa 921-8501, Japan  
b6400561@planet.kanazawa-it.ac.jp,  
tyama@neptune.kanazawa-it.ac.jp

**Abstract.** In recent years, interactive educational methods called “Active Learning” have often been introduced in educational institutions such as universities. However, active learning activities such as group work are sometimes difficult to evaluate because it is uncertain what types of outcome should be considered good or bad. To approach this problem, using sensors on a smartphone, we developed a support system visualizing situations such as group work activity or the degree of understanding of students. In this study, we focus on body movements that appear frequently in this group work and classify them to understand the group work situation more clearly. To classify body movements, we created a dataset consisting of 10 movements appearing in group work, such as “Nodding.” Using this dataset, we investigated whether the movements could be identified by the method of deep learning. As a result, it was found that body movements with few individual differences could be identified with relatively high accuracy.

**Keywords:** Active learning · Deep learning · Body movement · Communication · Smartphone · Human sensing

## 1 Introduction

In recent years, educational methods called “Active Learning” have been introduced in education, from primary schools to universities; students are supposed to learn more subjectively and interactively by this method. For example, in active learning, students often discuss a certain theme in groups and present the groups’ conclusions to other students. It is argued that this type of subjective activity has a higher educational value than the traditional “knowledge transfer” type of education [1, 2].

In an actual class, it may be difficult for teachers to evaluate whether group work is going well or whether students are acquiring the ability to communicate knowledge outside of a specific field. From this point of view, our research group developed a system to measure the group work situation using smartphone sensors so that teachers, as facilitators, can evaluate the activities of students and give better advice to students [3]. The developed system could speculate about body movements related to communication and the state of a dialogue situation [4]. This measurement system currently

only acquires values from an acceleration sensor, and the evaluation of the body movements is performed manually after the data acquisition. In the future, we will develop a system that can evaluate group work in real time using measured data. To realize such a system, it is necessary to identify body movements in real time and to evaluate group work from a time series of body movements.

Several previous studies have measured body movements in the education field. One of these studies analyzed the body movements of primary school students during classes using sensors like name tags [5]. In that research, relatively large movements such as “Standing upright and seating” and “Moving desk” were focused on and analyzed. In contrast, in the group work that we focus on, body movements related to communication such as “Nodding” or “Turning around” are considered essential; these have often been observed in our previous research [4]. In addition to these movements, casual body movements such as “Crossing legs” or “Touching face” have also often been confirmed in previous research [4]. If these various body movements appearing in group work can be identified in real time, the group work can be evaluated from the viewpoint of nonverbal communication, that constitutes most of human communication [6]. Therefore, in this research, we create datasets of body movements appearing in group work and investigate whether they can be identified by deep learning.

## 2 Methods

### 2.1 Body Movements Related to Group Work

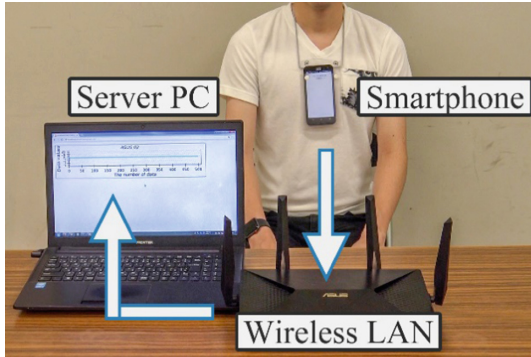
In this research, we investigated the recorded video from a previous study [4] again and selected 10 types of body movements to identify. Table 1 shows these movements. Identification numbers, M01–M10, are assigned to each of the 10 movements. M01–M05 were considered to be body movements related to communication, whereas M06–M10 were casual body movements appearing in group work.

**Table 1.** Ten types of body movements in group work.

Movement ID	Movement name
M01	Remaining stationary
M02	Clapping
M03	Raising hand
M04	Nodding
M05	Turning around
M06	Stretching
M07	Crossing legs
M08	Crossing arms
M09	Resting elbow on a desk
M10	Touching face

## 2.2 Data Measurement of Body Movements

Measurement of body movement data for deep learning was conducted with 11 university students (average age 21.7 years) using the system developed in the previous study [3] (Fig. 1). Measurement of five body movements were conducted as one session and the body movements related to communication were measured at first. Then, five casual body movements were measured.



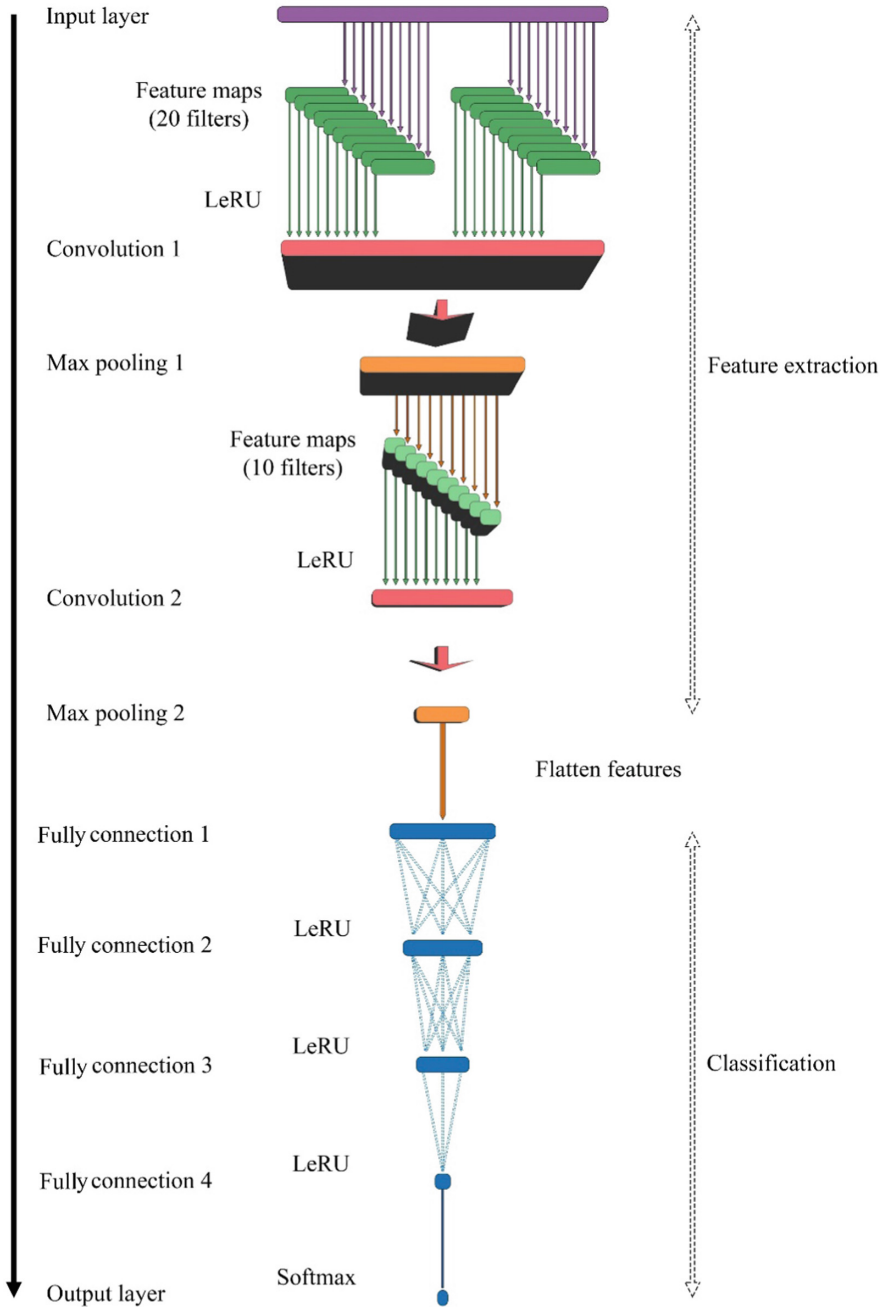
**Fig. 1.** Measurement system.

The measurement time was set to 1 min for each body movement, and each session lasted approximately 15 min including explanation and break time. Nine types of body movements—all except “Remaining stationary”—were repeated in accordance with the buzzer, which sounded at intervals of 5 s.

## 2.3 Preprocessing of Measured Data

In this research, the measured data from the acceleration sensor were used for learning using a convolutional neural network (CNN), which is one of the deep learning methods. In the CNN (Fig. 2), the following preprocessing was performed on the measured data.

1. Extract 4096 ms data (the norm of triaxial acceleration data) from 6000 ms data, which contain one body movement.
2. Apply Hanning window to the 4096 time series data and apply Fast Fourier Transform to the data.
3. Extract the power spectrum in the low frequency region (0–25 Hz) where the features of body movements appear.
4. The maximum value is found within each sample and normalization is performed to scale the input data to the range 0–1.



**Fig. 2.** Overview of CNN.

## 2.4 CNN for Deep Learning

The first step is to apply the preprocessing to all 110 data items (11 subjects  $\times$  10 movements). Next, labels are attached to the preprocessed data to pair the power spectrum data with the movement name, as the training data for one hot encoding. The created dataset is used for training a CNN, whose structure is shown in Fig. 2. An overview of the structure is as follows:

1. Convolution with 20 filters.
2. Max pooling with 1/8 size.
3. Convolution with 10 filters.
4. Max pooling with 1/2 size.
5. Fully connected layer with 90% dropout.

This research uses TensorFlow 1.10.0, which is one of the frameworks for deep learning, for CNN implementation.

## 2.5 Three Datasets for CNN

In order to investigate how the classification result changed depending on the learning data, we prepared three types of learning dataset:

1. Five types of body movement from M01–M05.
2. Five types of body movement from M06–M10.
3. Ten types of body movement from M01–M10.

To verify the learning results, 11 samples combining 10 items of training data and one item of test data (each sample contained 11 data items) were prepared in each dataset. In other words, we verified the CNN with learning data for 10 subjects by inputting one unknown subject.

# 3 Results

## 3.1 Classification of Body Movements

Table 2 shows all classification results for each dataset. For the results of learning the five movements M01–M05 (the body movements related to communication), the average accuracy is 87.64%. This means that it is possible to detect these body movements with relatively high accuracy by using the CNN. “Remaining stationary (M01),” “Clapping (M02),” and “Nodding (M04)” give better results (over 90%) than “Raising hand (M03)” and “Turning around (M05).”

As for the results of learning the five movements M06–M10 (the casual body movements appearing in group work), the average accuracy is 68.00%. Compared with M01–M05, it is more difficult to classify these five body movements by using the CNN. “Crossing legs (M07),” “Crossing arms (M08),” and “Resting elbow on a desk (M09)” give the best results. However, “Stretching (M06)” and “Touching face (M10)” are often classified as other movements.

**Table 2.** Results of classification using three datasets.

Subject	Accuracy for each dataset (%)		
	M01–M05	M06–M10	M01–M10
S01	84.00	84.00	68.00
S02	84.00	64.00	50.00
S03	80.00	68.00	56.00
S04	96.00	60.00	64.00
S05	92.00	72.00	62.00
S06	100.00	68.00	70.00
S07	88.00	72.00	68.00
S08	92.00	64.00	56.00
S09	80.00	52.00	50.00
S10	84.00	80.00	60.00
S11	84.00	64.00	66.00
Average	87.64	68.00	60.91

Finally, for the results of learning all 10 types of movement, the average accuracy is 60.91%. Of the three datasets, these types of data give the worst results. “Remaining stationary (M01),” “Clapping (M02),” and “Nodding (M04)” give better results than the other seven movements, which are often misclassified.

### 3.2 Verification of CNN Using Actual Group Work Data

The trained CNN was verified with the data about body movements in actual group work. For this, a real group work (20 min) involving three subjects was carried out three times; the three subjects learned one topic by teaching each other. In this research, movements by the nine subjects were identified with the learned CNN. Figure 3 shows a picture of the actual group work.

**Fig. 3.** Sample of actual group work.

Within the three 20-min group work sessions, the section where the discussion was held was extracted and this section was divided into 10-s segments. Next, the body

movements of each subject were visually inspected and labeled by a human analyzer, using the labels for the 10 types of body movements (M01–M10). As a result, 440 movements in the group work sessions were labeled. Table 3 shows the number of occurrences of each body movement.

**Table 3.** Number of movements that appeared in actual group work.

Movement ID	Number of samples
M01	32
M02	0
M03	0
M04	133
M05	18
M06	32
M07	4
M08	6
M09	53
M10	162

For verification, body movements in the actual group work were identified by the trained CNN for every 10-s segment. The CNN trained with the data of M01–M05 was used for the identification of M01–M05. Similarly, the CNNs trained with the data of M06–M10 and M01–M10 were used for the identification of M06–M10 and M01–M10, respectively.

Table 4 shows the results of the five movements from M01–M05. In this case, the total accuracy is 75.41% (138/183). This means that it is possible to detect the body movement of actual group work with high accuracy.

**Table 4.** Results of classification of movements M01–M05.

Input	M01	M02	M03	M04	M05
M01	<u>19</u>	0	0	13	0
M04	14	0	0	<u>119</u>	0
M05	8	0	0	10	<u>0</u>

Table 5 shows the results of M06–M10. In this case, the total accuracy is 63.81% (164/257). As with the test data, these five body movements are more difficult to classify in actual group work. In particular, “Crossing legs (M07),” “Crossing arms (M08),” and “Resting elbow on a desk (M09)” are often misclassified as other body movements.

Finally, Table 6 shows the results of body movements M01–M10. The accuracy was 46.82% (206/440). In this case, the answer rate is the worst, as with the test data.

Among these movements, only “Nodding (M04)” and “Touching face (M10)” give the best results, but even these movements are often misclassified.

**Table 5.** Results of classification of movements M06–M10.

Input	M06	M07	M08	M09	M10
M06	<u>10</u>	0	0	0	22
M07	1	<u>0</u>	0	0	3
M08	0	0	<u>0</u>	0	6
M09	4	0	0	<u>0</u>	49
M10	8	0	0	0	<u>154</u>

**Table 6.** Results of classification in movement M01–M10

Input	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10
M01	<u>5</u>	0	0	11	0	0	0	0	0	16
M04	1	0	0	<u>74</u>	0	0	0	0	0	58
M05	0	0	0	5	<u>0</u>	0	0	0	0	13
M06	0	0	0	14	0	<u>0</u>	0	0	3	15
M07	0	0	0	3	0	0	<u>0</u>	0	0	1
M08	0	0	0	3	0	0	0	<u>0</u>	0	3
M09	1	0	0	14	0	0	0	0	<u>0</u>	38
M10	2	0	0	33	0	0	0	0	0	<u>127</u>

## 4 Discussion

The results showed that the accuracy, for data consisting of five body movements related to communication, was 87.64% with test data and 75.41% with real data. This means that analyzing the frequency of acceleration data and learning its distribution with the CNN make it possible to identify body movements related to communication in group work with high accuracy.

Meanwhile, the accuracy for five casual body movements was 68.00% with test data and 63.81% with real data, which is more than 10% lower than the movements related to communication. This difference is caused by the individual differences in the casual body movements. For example, the “Stretching” movement was different depending on the subject, and such data would dirty the learning data for CNN. In contrast, “Clapping” and “Raising hand” varied in rhythm and speed depending on the subject, but the basic movement did not change between subjects. This explains why they could be learned by the CNN.

In this research, the dataset consisted of only 11 people, but if the dataset is enlarged, it is possible that the individual differences can be canceled. However, if only the quantity of training data is increased, overfitting may occur and the classification rate cannot be improved. To solve this problem, ensemble learning would be a better approach. For example, in the case of preparing data on 40 people, it is better to create



10-person data for one classifier and to prepare four weak classifiers than to prepare one classifier using all 40-person data. Using this method, even if there are data groups with a large amount of noise, its influence can be reduced. In addition, when increasing the quantity of data, it is possible to grow existing classifiers only by increasing their number.

## 5 Conclusions

In this research, to identify the body movements in group work for understanding the situation of active learning, the datasets of body movements appearing in group work were created and classified by deep learning. It was found that data groups composed of body movements with little individual differences can be identified with relatively high accuracy.

In future work, further learning data will be added while using an ensemble learning method, as explained above. At the same time, we will identify the body movements in more samples of actual group work. In this case, it is not clear whether the body movements used in this research necessarily yield good results. Therefore, we will further investigate the dataset of body movements and the structure of the classifier that can yield the best results. Furthermore, we will develop a system to evaluate the group work in real time using data classified by CNNs so that facilitators can realize better active learning class.

**Acknowledgments.** This work was supported by JSPS KAKENHI Grant Numbers 16H0322500 and 18K1141200.

## References

1. Dale, E.: *Audio-Visual Method in Teaching*. The Dryden Press, New York (1946)
2. Crawley, F.E., Malmqvist, J., Östlund, S., Brodeur, R.D., Edstrom, K.: *Rethinking Engineering Education: The CDIO Approach*. Springer, London (2014). <https://doi.org/10.1007/978-0-387-38290-6>
3. Kimura, M., Kato, Y., Hagiwara, T., Yamamoto, T., Miyake, Y.: Measuring system of communication between a lecturer and students in a class. *Correspondences Hum. Interface* **18**(2), 113–118 (2016)
4. Harada, N., Kimura, M., Yamamoto, T., Miyake, Y.: System for measuring teacher–student communication in the classroom using smartphone accelerometer sensors. In: Kurosu, M. (ed.) *HCI 2017, Part II. LNCS*, vol. 10272, pp. 309–318. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-58077-7\\_24](https://doi.org/10.1007/978-3-319-58077-7_24)
5. Yamamori, K., Ito, T., Nakamoto, K., Haggiwara, Y., Tokuoka, M., Oouchi, Y.: Recording elementary school students’ academic engagement or on-task behavior using accelerometers. *Jpn. J. Educ. Technol.* **41**(2), 501–510 (2018)
6. Birdwhistell, R.L.: *Kinesics and Context: Essays on Body Motion Communication*. University of Pennsylvania Press (1970)