Observing Facial Muscles to Estimate the Learning State During Collaborative Learning: A Focus on the ICAP Framework

Yuying CAI[a*], Shigen Shimojo[b] & Yugo Hayashi[a]
[a] *College of Comprehensive Psychology, Ritsumeikan University, Japan*
[b] *Graduate School of Human Science, Ritsumeikan University, Japan*
* cp0060ex@ed.ritsumei.ac.jp

Abstract: This study tests the proposal that a learner's learning state can be estimated by observing their facial expressions. This could prove helpful for adaptive learning applications using facial recognition technology to provide appropriate feedback to learners. Based on our knowledge of facial expressions and use of the Interactive, Constructive, and Active processes from the ICAP framework, we hypothesized the following: (1) During the Active process, because utterances consist of reading the learning material, the muscles of the mouth are primarily in motion. (2) During the Constructive process, the muscles of the mouth move, and the eyelids tighten, because self-reflection and thoughtful utterances are required. (3) During the Interactive process, the conversation involves conflict; therefore, the eyebrows raise and the eyes open. To that end, we recorded and analyzed the facial expressions and utterances of five pairs of learners during the three learning states. We then organized the data by facial muscles that appeared most frequently during the respective learning states. The analyses generally supported our hypotheses. However, micro-facial expressions generated by facial muscles other than those considered in our hypotheses are also relevant and should be explored in future research.

Keywords: Facial expression analysis, ICAP framework, collaborative learning, learning support, adaptive learning

1. Introduction

Studies have found that during collaboration, constructive interaction that reconstructs one's understanding from different perspectives is important (Chi, De Leeuw, Chiu & Lavancher, 1994; Okada & Simon, 1997; Shirouzu, Miyake & Masukawa, 2002). In recent years, automatic estimation of the collaborative learning process has attracted attention in the field of educational technology. During collaborative learning, it is important to support learners based on their state of learning (Hayashi, 2019a). In this study, we focus on the learner's facial expressions during collaborative learning and examine whether they are effective indicators for estimating a learner's state.

In Section 1.1, we explain the requirements for collaborative learning support. In Section 1.2, we discuss the research on learning support using facial expressions and explain the relationship between facial expressions and the collaborative process, which is the focus of this study. Then, we state our purposes and hypotheses (Section 1.3).

1.1 Support in collaborative learning

Researchers have confirmed that learning and performance are both promoted by interacting with others in cooperation (Chi, De Leeuw, Chiu & Lavancher, 1994; Okada & Simon, 1997; Shirouzu, Miyake & Masukawa, 2002). During tasks, learning in pairs is better than individual learning for making new discoveries, facilitating explanatory activities, and participating actively, resulting in high performance (Okada & Simon, 1997). Additionally, self-explanation during collaborative problem solving integrates existing knowledge with new knowledge (Chi, De Leeuw, Chiu & Lavancher, 1994). Furthermore, constructive interaction that reconstructs one's understanding based on different viewpoints in

collaborative learning is important in certain tasks, such as origami (Shirouzu, Miyake & Masukawa, 2002). Jigsaw learning research is an example of practical research that focuses on cooperative activities based on different perspectives. With the jigsaw learning method, students learn by bringing complementary knowledge to help solve a problem or complete a task. For example, Zhu & Sunakawa (2019) analyzed the interview data of students who took a class taught by the jigsaw learning method and examined the factors that caused many students to change from the "self-study type" to the "collaboration type." As an example of laboratory research, Hayashi, Miwa & Morita (2007) examined cooperative learning based on different viewpoints using an experimental task that uses the principle of illusion, but found that it was difficult to obtain the viewpoint of both learners in the dyad.

As described above, it is important to reconstruct one's understanding based on the different viewpoints of others. However, researchers have found that it is difficult to eliminate communication discrepancies during such learning activities, and that they can result in failure to acquire correct perspectives of others and constructive interaction (Hayashi, Miwa & Morita, 2007). Therefore, when creating a support system for collaborative learning, it is necessary to observe the collaborative process between learners and support it.

One of the tasks involved in developing a support system for collaborative learning is providing relevant feedback based on the state of the learner. This type of support is called adaptive feedback. An example of an adaptive feedback system is the Cognitive Tutor (Anderson, Corbett, Koedinger & Pelletier, 1995), which detects the state of the learner using a model and then provides relevant feedback. However, few studies have examined adaptive feedback in collaborative learning. Detecting learner states during the collaborative learning process, and providing them with appropriate feedback during such states, are research themes that are worthy of attention. For example, in a study using pedagogical conversation agents (like online assistants) to facilitate collaborative learning, Hayashi (2019b) established the effectiveness of estimating the collaborative process based on two indicators, gaze recurrence and language alignment. However, there is need to increase the number of indicators for detecting the state of the learner during collaborative learning. Therefore, in this study, we focus on facial expressions and non-verbal information, which are important during communication.

1.2 Detecting learner state using facial expression analysis

Facial expression research has been conducted for many years, and it is historically known that facial expressions can be used to estimate emotions (Ekman & Ancoli, 1980). The Facial Action Coding System (FACS) is a comprehensive system for distinguishing facial movements Ekman & Freisen (1976) using facial expression muscles, described as action units (AUs). For example, the emotion *worry* is related to AU04 (Brow Lowerer) and AU07 (Lid Tightener). In another study using FACS, the relationship between the learner's facial expressions and emotional state was examined: facial expressions could help estimate emotions related to mutual understanding during collaborative learning (Hayashi, 2019b). These studies suggest that the learner state can be detected from the learner's facial expression. However, although we have investigated relationships between facial expressions and emotional states, it is necessary to investigate the relationship between facial expressions and the learning process from the context of providing adaptive feedback based on the learner's state.

The ICAP framework classifies learning activities into Passive, Active, Constructive, and Interactive (Chi & Wylie, 2014; Chi, 2009). In the present study, we examine the learner's activity state (Active, Constructive, and Interactive) during collaborative learning to capture the deep interaction of the collaboration (Chi & Wylie, 2014). The relationship between facial expressions and emotional states is not analyzed here.

Now let us look at the ICAP modes in detail. In the passive mode, it can be assumed that the knowledge change process is isolated in an encapsulated manner to "store" the received information. A learner directly collects information from learning materials during lessons, and there is no other behavior related to learning; for example, listening to a lecture without taking notes. In the active mode, the learner is not required to act, such as when reading learning materials. Once the relevant schema is activated, the learner can absorb or integrate new information into the activated schema, so that the learner can fill in the gaps in the model and make it more complete. In the constructive mode, the learner delves into the learning material and externalizes the content. Constructive behaviors often require the processes of "inferring" and inferring means a process of elaborating, such as adding more details or qualifications. For example, if you learn A and B and associate them, you then possess an utterance

about their contents and have created a link on the concept map. In the Interactive mode, a learner constructs thoughts based on what someone else has said; having doubts about the content; objecting to the content; or having a conversation involving conflicts. Each member of the dyad must be constructive, thereby engaging in the cognitive processes of activating, integrating, and inferring. For example, a learner finds a contradiction in the concept that the other learner explained and requests them to explain it further. Based on the above, when FACS is used to predict the learner's ICAP modes, the following will be true. In the Active mode, since the utterance is simply a reading of the learning material, it is mainly the muscles of the mouth which move. In the Constructive mode, since the utterance is engaged in self-reflection and results from thinking deeply, in addition to the above described movement of the muscles of the mouth, the eyelids tighten. In the Interactive mode, since the utterance is in a conversation involving conflicts, other facial muscles come into play and movements such as raising of the eyebrows and opening of the eyes are observable.

1.3 Purpose and hypothesis

To support the learner during collaborative learning, it is necessary to make a human facial expression model that can estimate the state of the learner during the collaborative process. Therefore, this study has two purposes. The first is to estimate which ICAP learning process a learner is engaged in based on their AUs. The second is to sort the types of AUs present in each ICAP modes. We will verify the following hypotheses.

H1-1: In the Active mode, since the utterance is simply a reading of the learning material, the main facial movement is from the mouth muscles, such as those are used to making dimples.

H1-2: In the Constructive mode, since the utterance is a result of self-reflection and thinking deeply, in addition to the muscles of the mouth, the eyelids tighten.

H1-3: In the Interactive mode, since the utterance is produced in a conversation involving conflicts, we can observe the movement of facial muscles such as those needed for raising the eyebrows and opening the eyes.

2. Method

The present study used the ten people from Shimojo & Hayashi (submitted) as a sample. Here, pairs of learners worked on a collaborative learning task that required explaining a specific psychological phenomenon while creating a concept map. For details, refer to (Shimojo & Hayashi, 2019a; Shimojo & Hayashi, 2019b; Shimojo & Hayashi, 2019c).

2.1 Participants analyzed in this study

We analyzed ten Japanese university students (1 male and 9 female) which had successful collaborative process. Two were excluded because they did not meet the criteria for classification within the ICAP framework.

2.2 Materials and systems

Two PCs and two monitors were used by the participants, and two Sony HDRCX680 video recorders filmed their conversations and facial expressions. We also used Cmap Tools, a tool for creating concept maps (https://cmap.ihmc.us/). A monitor and video recorder were placed in front of each participant in a pair. They sat across to each other and a partition was put between them so they cannot see each other.

2.3 Procedure

In the experiment, the task was to make inferences about a certain psychological phenomenon. Before the task, participants read a text about causal attribution. Then, there was a story that participants participate in school counseling with Michael Peter and he is worried about the new semester, they need

to explain why the student is worried about the new semester based on the story by causal attribution. Furthermore, the concept map was used as a tool for explaining activities, and the facial expression data during the conversation were collected.

In this study, based on the data in Shimojo & Hayashi (submitted), we analyzed the videos of the participants during the Interactive, Constructive, and Active modes; we verified their learning state by their utterances. We analyzed the video using OpenFace software.

2.4 Dependent variable: facial expression analysis

The utterances of the participants were classified using the ICAP framework, and the facial movements they demonstrated were analyzed by OpenFace in 1/20 second units. OpenFace automatically calculated whether an AU appeared. The numerical value output by OpenFace indicated the strength of the AU and was obtained by a formula described in a "toolkit" for using the software (Baltrušaitis, Robinson & Morency, 2016). There were 18 types of AUs observed among the participants as follows. AU01: Inner Brow Raiser, AU02: Outer Brow Raiser, AU04: Brow Lowerer, AU05: Upper Lid Raiser, AU06: Cheek Raiser, AU07: Lid Tightener, AU09: Nose Wrinkler, AU10: Upper Lip Raiser, AU12: Lip Corner Puller, AU14: Dimpler, AU15: Lip Corner Depressor, AU17: Chin Raiser, AU20: Lip stretcher, AU23: Lip Tightener, AU25: Lips part, AU26: Jaw Drop, AU28: Lip Suck, and AU45: Blink.

After determining the AUs involved, we analyzed the video of the utterances for each ICAP learning process. The frequency of each AU was tabulated for each video (second). In the analysis, the Z score was calculated. The video comprises 418.72 minutes of the Interactive mode, 158.19 minutes of the Constructive mode, and 123.85 minutes of the Active mode.

3. Result

3.1 Checking the reliability of automatic coding

Before verifying the hypothesis, we confirmed the reliability of the AUs calculated by OpenFace. We analyzed 6 of the 18 AUs by sampling all participants' data. Then, we set the coding standard for each AU facial expression, and performed manual labeling. For example, the coding standard for AU45 was defined as the movement of "closing and opening eyes."

To calculate the degree of agreement between the manual and the OpenFace coding results, we converted the data to obtain the Z score. We analyzed human detection of AUs against OpenFace detection of AUs. Table 1 demonstrates the overall concordance rate. From this result, Table 1 shows that the degree of agreement regarding AU07 (Lid Tightener) is low, but that the degree of agreement regarding the other AUs is high. The analysis revealed that there was the correlation of Z score between manual coding and OpenFace coding. Based on these results, we attempted to confirm H1-1, H1-2, and H1-3 using OpenFace coding.

Table 1. *Concordance rate by AU for analysis by human and OpenFace*

Action Unit	Interactive	Constructive	Active
AU02 (Outer Brow Raiser)	87%	90%	85%
AU05 (Upper Lid Raiser)	100%	55%	95%
AU07 (Lid Tightener)	12%	6%	29%
AU17 (Chin Raiser)	76%	99%	78%
AU20 (Lip stretcher)	98%	100%	95%
AU45 (Blink)	92%	95%	95%

3.2 Analysis of the relationship between ICAP and AUs

In this section, we describe the results of the OpenFace analysis and show which AUs appeared during each mode of the ICAP process. Figure 1 shows the Z score of each AU for each ICAP mode, and the details of the analysis results are described below.
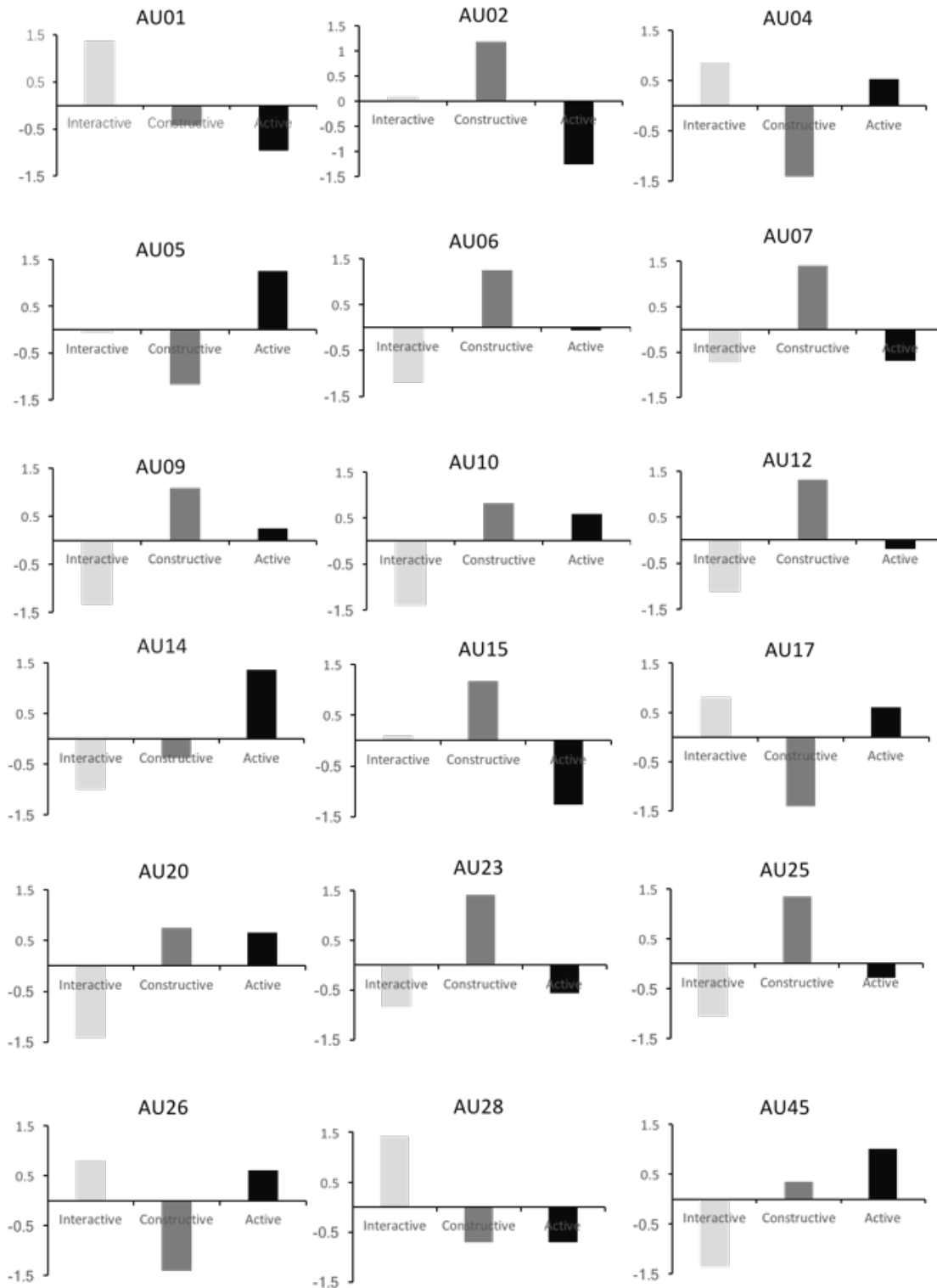
Figure 1. Z score of each AU for ICA.

3.2.1 AUs that frequently appear in the Active mode

In the Active mode, the frequency of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink) is high. The Z score was calculated, and a score higher than 1 was judged to be high in frequency. The same standard was applied to the Constructive and Interactive modes. We also examined the difference between AUs higher than 1. To examine which facial expression muscle appeared most strongly, we compared the average values of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink).

Therefore, a one-factor within-participant analysis of variance was performed and no significant difference was confirmed (F (2, 14) = 1.80, p = .20, η_p^2 = .21). All three AUs appeared at the same level.

From the above, it can be said that H1-1 was generally supported because AU14(Dimpler) appeared at a high frequency during the Active mode.

3.2.2 AUs that frequently appear in the Constructive mode

In the Constructive mode, we found that the frequency of AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part) was high. Then, the average values of AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part) with Z score higher than 1 were compared. A one-factor experiment within-participant analysis of variance confirmed a significant difference (F (7, 49) = 3.48, p <.01). Therefore, when multiple comparisons were performed to confirm differences, no significant difference was observed. These AUs appeared to the same extent.

Based on the above, the frequency of AU07 (Lid Tightener) was high and the frequency of AU14 was not high, but the frequency of AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part), which are the movements of the mouth muscles, were high.

3.2.3 AUs that frequently appear in the Interactive mode

By the same criteria used for the Active and Constructive modes, we found that AU01 (Inner Brow Raiser) and AU28 (Lip Suck) had a high frequency in the Interactive mode. Then, we compared the average values of AU01 (Inner Brow Raiser) and AU28 (Lip Suck) with Z scores higher than 1. A paired t-test revealed that AU01 (Inner Brow Raiser) was significantly higher than AU28 (Lip Suck) (t (7) = 4.12, p <.01).

Therefore, the frequency of AU05 (Upper Lid Raiser) is not high, but the frequency of AU01 (Inner Brow Raiser) is high. Therefore, it can be said that H1-3 was partially supported.

3.2.4 Concluding remarks on AUs that appear frequently

The statistical analysis clarified that different AUs appear in each mode of ICAP's collaborative learning process. The study's second purpose was to categorize AU types by their occurrence during each ICAP mode, as summarized in Table 2. The order of the type of the AU shown in the table occurrence frequency is placed in the order by the type of AU with high frequency.

From this we determined that there are no AUs that overlap the different ICAP modes; as such, this method is useful for estimating which ICAP mode the learner is experiencing. A more detailed discussion of this point will be provided in the upcoming section.

Table 2. *AUs that have a high frequency of occurrence by ICA*

ICAP framework	AU that frequently appear
Interactive	AU01 (Inner Brow Raiser), AU28 (Lip Suck)
Constructive	AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener), AU25 (Lips part)
Active	AU05 (Upper Lid Raiser), AU14 (Dimpler), AU45 (Blink)

4. Discussion

In H1-1 we predicted that the muscles that move the mouth, such as when making dimples, were the main type of AUs present during the Active mode. In H1-2, we predicted that during the Constructive mode, in addition to the muscles of the moving mouth, the eyelids would tighten. In H1-3, we predicted frequent eyebrow raising and eye opening during the Interactive mode. Our examination revealed that the frequency of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink) was high during the Active mode; AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), and AU07 (Lid Tightener) during the Constructive mode; and AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener), AU25 (Lips part), AU01 (Inner Brow Raiser), and AU28 (Lip Suck) during the Interactive mode. H1-1 is generally supported by these results because of the high frequency of AU14 (Dimpler); H1-2 is generally supported because of the high frequency of AU07 (Lid Tightener), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part); and H1-3 is partially supported because of the high frequency of AU01 (Inner Brow Raiser). In addition, we found that different high frequency AUs appeared in each ICAP phase. As there is no overlap of high frequency AUs between phases, observing AUs can be considered useful for estimating ICAP. For example, if AU01 (Inner Brow Raiser) and AU28 (Lip Suck) both appeared on the face of the learner, there is a high possibility that the learner is in the Interactive mode. Furthermore, results show that micro-facial expressions produced by facial muscles that were not mentioned in the hypotheses were additionally found. For example, the frequency of AU09(Nose Wrinkler) was a type of muscles that we did not consider in our hypotheses, but was detected during the Constructive mode.

Previous studies have attempted to estimate emotions by analyzing facial expressions. For example, researchers have clarified that facial expressions can be used to infer emotions related to mutual understanding in collaborative learning (Hayashi, 2019b). However, there has been little research on estimating the learner's mode during the collaborative process, which is required for adaptive feedback. Moreover, this is the first attempt to classify facial expressions performed during the ICAP modes proposed in Chi and Wylie's work on the topic; they categorized the ICAP mode of the learner by utterance and behavior (Chi & Wylie, 2014). Because the AUs with a high frequency of appearance differed for each ICAP mode, AU can estimate the learning process of ICA. Thereby, providing adaptive feedback based on the learner's learning states will be possible. So, this study provides important knowledge for effective adaptive support in future collaborative learning models.

In addition, ICAP is classified by utterances and actions (Chi & Wylie, 2014). In this study, we analyzed the data classified into ICAP based on the learner's conversational activities. Therefore, it is necessary to more strictly classify into ICAP by not only the conversational activities but also the indicators such as learner's behavior and gazes. ICAP predicts different learning levels, Active realizes a higher learning level than Passive, Constructive realizes a higher learning level than Active, and Interactive realizes a higher learning level than Constructive (Chi & Wylie, 2014). In this study, we examined AUs that have a high frequency of occurrence for each ICA, but we did not consider Passive. Since the learner's state needs to be promoted the transition from Passive to Active, it is necessary to provide adaptive feedback when detecting Passive. Therefore, it is considered that the passive feedback can be estimated from the AU and adaptive feedback can be performed by examining the AU that frequently appears in the passive.

5. Conclusion

In this study, we explored how learners can be better supported during collaborative learning by quantitatively investigating learners' learning process. To that end, we analyzed facial expressions during different ICAP modes by observing the movement of their AUs during the aforementioned states. In the present study, (1) we found that the learning state can be estimated by observing the learner's facial muscles, and (2) we organized the relevant AUs by their frequency during the respective ICAP modes. We used utterance and facial expression data obtained from pairs of learners while they performed a collaborative learning concept explanation task. The results of the data analysis generally supported our three hypotheses. On the other hand, it became clear that micro-facial expressions produced by facial muscles other than those considered in the hypotheses are also related. In future

studies based on the AUs classified in this study, we will examine their relationship with the learning process in more detail and perform deeper quantitative analysis to fine tune the development of effective adaptive feedback models.

Acknowledgements

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167-207.

Baltrušaitis, T., Robinson, P. & Morency, L. (2016) OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision*, 1–10.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Cognitive Science*, *1*(1), 73-105.

Chi, M. T. H, Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219-243.

Ekman, P. & Ancoli, S. (1980). Facial Signs of Emotional Experience. *Journal of Personality and Social Psychology*, *39*(6), 1125-1134.

Ekman, P. & Freisen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, *1*(1), 56-75.

Hayashi, Y. (2019a). Towards Supporting Collaborative Learning with an Intelligent Tutoring System: Predicting Learning Process by Using Gaze and Verbal Information. *Cognitive Studies, 26*(3), 343-356.

Hayashi, Y. (2019b). Detecting collaborative learning through emotions: An investigation using facial expression recognition. *Proceedings of the 15th International Conference on Intelligent Tutoring Systems (ITS2019)*, 89-98.

Hayashi, Y., Miwa, K. & Morita, J. (2007). A Laboratory Study on Collaborative Problem Solving by Taking Different Perspectives. *Cognitive Studies, 14*(4), 604-619.

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science, 21*(2), 109-146.

Shimojo, S. & Hayashi, Y. (2019a). An experimental investigation on collaborative dyads' explanation activities using conceptual maps: Analysis on learning performance based on understanding and the use of different perspectives. *IEICE technical report, 119*(39), 87-91.

Shimojo, S. & Hayashi, Y. (2019b). Relation between dialog activity and learning performance on collaborative learning visualized other knowledge: An analysis of turn-taking and knowledge convergence. *Japanese Cognitive Science Society, 36*, 2-46

Shimojo, S. & Hayashi, Y. (2019c). How shared concept mapping facilitates explanation activities in collaborative learning: An experimental investigation into learning performance in the context of different perspectives. *Proceedings of the 27th International Conference on Computers in Education(ICCE2019)*, 172-177.

Shimojo, S. & Hayashi, Y. (submitted).

Shirouzu, H., Miyake, N. & Masukawa, H. (2002). Cognitively active externalization for situated reflection. *Cognitive Science, 26*(4), 469-501.

Zhu, G. & Sunakawa, Y. (2019). The Effect of the Jigsaw Learning Method on Students' Attitudes toward Their Academic Performance: Coordination between Activities. *The Society for Teaching Japanese as a Foreign Language, 145*, 25-36.