Multimodal Interaction Aware Platform for Collaborative Learning

Aoi SUGIMOTO*, **Yuki HAYASHI & Kazuhisa SETA**
Graduate School of Humanities and Sustainable System Sciences,
Osaka Prefecture University, Japan
*sugimoto@ksm.kis.osakafu-u.ac.jp

Abstract: A number of studies in the research field of Computer Supported Collaborative Learning (CSCL) have proposed various systems in order to facilitate learning in the context of social interactions. In a collaborative learning, participants exchange not only verbal but also nonverbal cues such as utterance, gaze and gesture for maintaining the relationships among one another. Nevertheless, very few attempts have been made to construct a CSCL system that can utilize such multimodal (verbal and non-verbal) information to support communication in a collaborative learning. In this paper, we propose a novel platform that enables CSCL system developers to construct their learning support tools that have original functions to process such multimodal information. By building learning support tools on our multimodal aware platform, we confirmed its usefulness and also potential to pioneer unexploited filed of research in learning analytics for CSCL and methods to intervene in collaborative learning processes using verbal and non-verbal information unutilized so far.

Keywords: CSCL, Multiparty Multimodal Interaction, System Development Platform, Verbal and Non-verbal Information

1. Introduction

In our globalized cross-border world, we are required to cultivate social interaction skills that enable us to collaboratively make decisions or solve problems with other colleagues in various situations (Griffin, et al., 2012). In order to cultivate such social interaction skills, meaningfulness of collaborative learning style based on social constructivism whereby plural participants acquire knowledge and solve a problem is widely recognized. For a successful collaborative learning whereby all the participants can get fruitful learning outcomes, participants' mutual engagements to the learning processes are required in addition to solving the problem itself. Participants get several benefits from such interactions, going from constructing deeper level learning, shared understanding, to developing social and communication skills and so on (Kreijns, 2003). Nevertheless, the quality of learning effects is not always assured due to the negative aspects of small group interactions, such as social pressure, inter- and intragroup aggression or conflict and polarization (Strijbos, 2011). In the research field of Computer Supported Collaborative Learning (CSCL), a number of CSCL systems have been proposed for supporting the learning processes using information communication technologies (Jermann, et al., 2001).

On the other hand, many studies on analyzing the small-group face-to-face interactions have been conducted in the research field of multiparty multimodal interaction (Gatica-Perez, 2011). In these studies, interaction management, internal states, social relationships, and so on have been analyzed and modeled based on integrating information via multimodal verbal and non-verbal communication channels such as utterance, gaze and gesture. These findings demonstrate the potential to develop novel CSCL systems that can analyze, assess and also intervene in various learning situations in real time.

However, there is no practical CSCL system embedding these findings in the field of multiparty multimodal interaction. One of the reasons underlying such situation is the lack of applicable platform for developing CSCL systems that can deal with various verbal and non-verbal information (multimodal information).

In this study, we aim to propose a verbal and non-verbal aware platform for developing CSCL systems. The proposed platform is intended to equip a fundamental infrastructure required for any CSCL systems, e.g., session management, and allow developers to implement/extend learning support

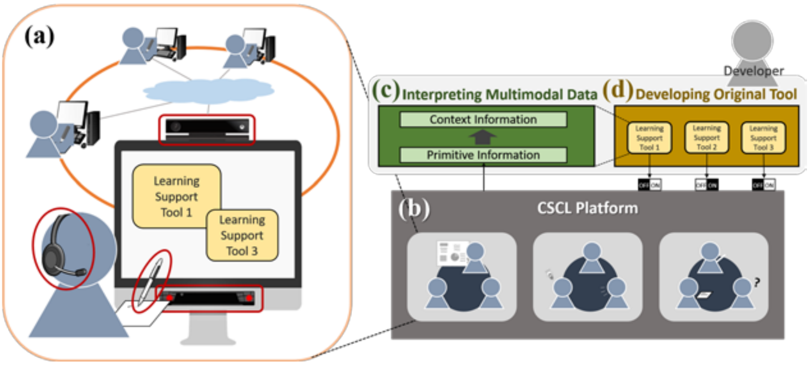


Figure 1. Platform Concept towards Developing CSCL Systems.

tools that can handle nonverbal as well as verbal information provided by the platform to facilitate fruitful communication during collaborative learning processes.

This paper is structured as follows: in section 2, we specify the requirements for developing the verbal and non-verbal aware platform; in section 3, we explain the architecture of the platform with its design principle; in section 4, we discuss the usefulness of the platform by showing an example of learning support tool developed on the platform; in section 5, we introduce some related works and argue the potential of the platform as a verbal and nonverbal aware CSCL system development environment.

2. Requirements

Figure 1 overviews our platform for CSCL systems development. This platform is equipped with various sensing devices in order to capture several verbal and non-verbal information of participants in a collaborative learning (Fig.1(a)). It also provides a fundamental infrastructure for developing CSCL systems, i.e., network and session management (Fig.1(b)). Furthermore, it provides a framework, which allows developers to specify rules to interpret sensed verbal and non-verbal 'primitive' information into 'context' information among collaborative learners (Fig.1(c)). Consequently, CSCL system developers can concentrate on developing their learning support tools with multimodal interpretation processing (Fig.1(d)) as well as specifying interpretation rules without getting involved in time consuming work for implementing lower level processing.
In order to realize the platform, the following two major requirements must be satisfied:

> *R1: A mechanism to provide several kinds of primitive verbal and non-verbal information which is the basis for multimodal interpretation (context information).*
> *R2: A mechanism for developers to define learning support tool specific information types (message types), and properly make them communicate in parallel.*

2.1 Requirement for Multimodal Interpretation Processing

In order to facilitate the analysis and understanding of conversational structures in multiparty multimodal interaction, Sumi et al. (2011) proposed a layered analysis model as shown in Table 1. The model represents four types of layers combining simple verbal and non-verbal communication signals exchanged among participants in order to achieve multimodal interpretation processing that elicits contextual information such as dominant level transition or participants' motivation towards their

Table 1: Layered analysis model for human interaction (Sumi et al., 2011).

Layer	Summary	Example
Interaction Context	the flow of interaction	dominant level transition
Interaction Event	the combinations of multiple primitive data	joint attention
Interaction Primitive	a single motion by a human	looking, speaking

interactions. According to the model, it is first necessary to capture raw data series such as voice, head-movement and eye-coordinates data (*raw data layer*). From the raw data series, interaction primitive elements such as who is speaking, who is gazing at whom and who is writing (*interaction primitive layer*) are extracted. The combination of interaction primitive elements leads to identifying interaction events such as joint attention or mutual gaze (*interaction event layer*). Based on the interaction events, one may infer higher-level contextual interpretation (*interaction context layer*).

In this research, we take the model as a grain size of layered concept for multimodal interpretation. In order for CSCL developers to build a system which equips such multimodal interpretation mechanism, it is required to pursue a stepwise processing beginning with designing learning activities, preparing an environment for detecting raw data from several sensing devices, extracting interaction primitives from the data, and interpreting them as interaction events.

In this research, in order to reduce CSCL developers' huge amount of workload, we propose a platform that allows them to access multimodal information without having to care of implementing the detection processing. Here, it is notable that the interaction elements focused on by developers depend on the nature of the collaborative learning and the learning subjects, and the way to deal with the detected data varies according to learning support tools. With keeping in mind this, our platform provides primitive information corresponding to raw data and interaction primitive layers in Table 1, which can be considered as the basis for multimodal interpretation (*R1*).

2.2 Requirement for Developing Learning Support Tools

There are many types of learning support tools used in collaborative learning, e.g., video-chat tool and text-chat tool as a means of communication, web browser tool for gathering information and shared-board tool for graphical representations, slideshow tool for presentations, etc. In addition to these tools, developers might have to develop their own specific learning support tools according to the target learning activity and subject.

In order to make developed learning support tools run in a network environment on the platform, it is necessary to equip the platform with a communication mechanism which handles various types of messages including learning support tools' specific ones during the learning activity. For example, the platform needs a specification about how to handle sending/receiving of each message from the learning support tool, meaning that input text messages in the case of text-chat tool, and drawing coordinates messages in the case of shared-board tool for instance, need to be properly specified. Hence, the platform should have a mechanism to communicate not only pre-specified information for authentication and raw level multimodal information captured by sensing devices, but also learning support tools' specific messages defined by developers (*R2*).

3. Platform for Developing CSCL Systems

3.1 Platform Architecture

Figure 2 illustrates the architecture of our platform. We employ a client-server architecture style to connect learning support tools used in a collaborative learning session. In the platform, *message communication modules* in client and server side function in synergy to ensure an adequate distribution of data to the requesting learning support tools.

On the server side, a relational database (*CSCL Database*) is equipped to store and manage the information about users, sessions, learning history, etc. In addition, to deal with audio and video streaming, the server is implemented by extending Red5 media server (Red5), which supports the real-time messaging protocol (RTMP). *Stream communication module* distributes audio and video data to the requesting learning support tools. *Session management module* manages participants' status and their active learning sessions. *CL-data management module* registers users' verbal and non-verbal information and their learning logs sent from clients to CSCL Database.

On the client side, *user management module* performs authentication processes by communicating with s*ession management module* on the server side. *Multimodal information management module* processes the data stream of the equipped sensing devices in a timely manner, and sends the data to the server. *Learning support tool management module* manages a group of learning

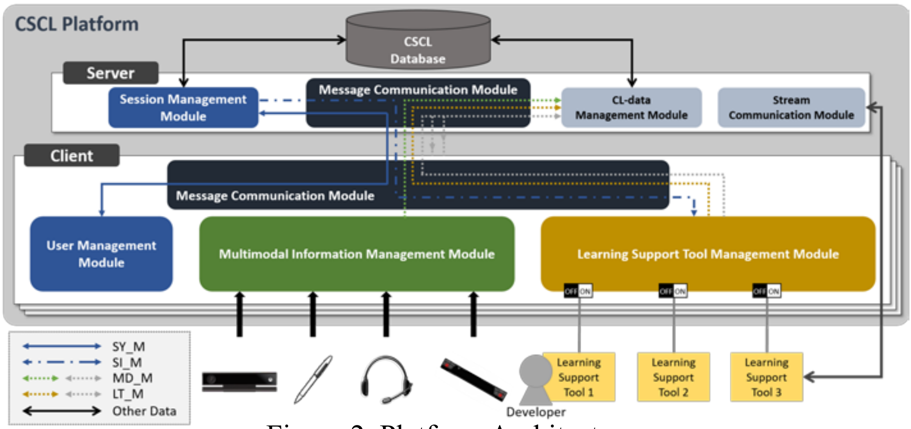

CSCL Platform
CSCL Database
Server
Session Management Module
Message Communication Module
CL-data Management Module
Stream Communication Module
Client
Message Communication Module
User Management Module
Multimodal Information Management Module
Learning Support Tool Management Module
OFF ON
OFF ON
OFF ON
SY_M
SI_M
MD_M
LT_M
Other Data
Developer
Learning Support Tool 1
Learning Support Tool 2
Learning Support Tool 3

Figure 2. Platform Architecture.

support tools used in active learning sessions, and distributes messages received from server to the tools properly.

Hereafter, we explain the core functions equipped on the platform in order to satisfy *R1&R2*; in section 3.2, we describe the mechanism that provides verbal and non-verbal information for system developers, and in section 3.3, we provide details about the mechanism that properly distributes messages data exchanged among the server and clients.

3.2 Verbal and Non-verbal Information Accessed by Developers

In order to satisfy *R1*, our platform provides developers with several types of verbal and non-verbal information as a basis for multimodal interpretation.

Currently, the platform provides four types of participants' behavioral information: *utterance*, *gaze*, *writing action* and *head movement*, each of which is often used as a feature to analyze a multiparty multimodal interaction, according to a survey article (Gatica-Perez, 2011) which reviews several topics in multimodal interaction research such as recognizing conversation structure based on speaker-addressee information, estimating roles in a conversation, and identifying change of dominance of a conversation group.

Table 2 summarizes verbal and non-verbal information provided by the platform. Each of the content in the column "Layer" represents the corresponding layer of multimodal interpretation model shown in Table 1. The grain size of the corresponding information is the individual behavior (*raw data* / *interaction primitive* in Table 1) of a certain participant. Developers are able to access the necessary information in order to interpret them into higher level of multimodal interaction.

(1) *Utterance: Speech interval* and the *content of utterance* are detected as utterance information. To capture them, a participant's utterance is recorded via a microphone device. The start and end of an utterance are detected when the audio level exceeds and falls below a certain threshold, respectively. Furthermore, as verbal information, the content of each utterance is provided using a speech-recognition API.

(2) *Gaze:* It is well known that the gaze interaction in communication or collaborative activity has several social functions such as expressing one's intention/feelings and regulating turn-taking (Kendon, 1967). In order to capture the participants' gaze information, our platform is developed on the premise of using screen-based eye-tracking devices. The platform provides a function for developers to set area-

Table 2: Accessible verbal and non-verbal information.

Type	Device	Information	Layer
utterance	microphone	speech interval	interaction primitive
		content of utterance	interaction primitive
Gaze	eye-tracker	eye-coordinates data	raw data
		gazing interval	interaction primitive
		target object	interaction primitive
writing action	digital pen	timing of writing	interaction primitive
head movement	depth camera	head direction data (roll, pitch, yaw)	raw data

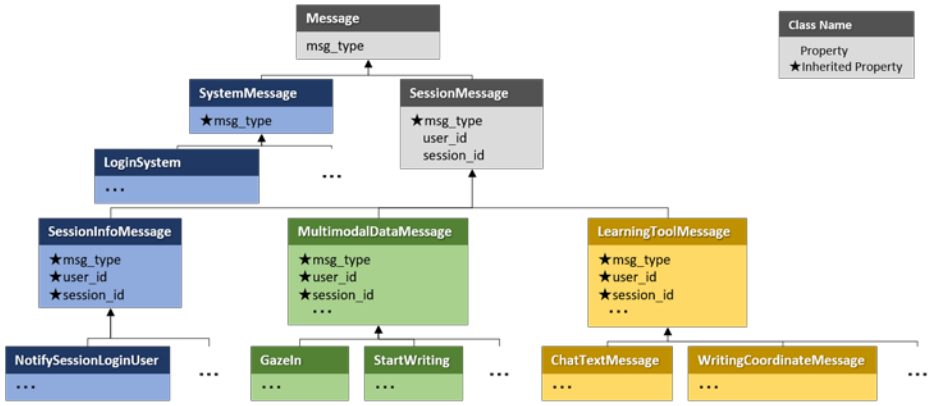


Figure 3. Hierarchical Message Class Structure.

of-interest (AOI) regions corresponding to each GUI unit such as learning support tools' window, video object in video-chat tool and text label in text-chat tool. Based on the registered AOI regions, it judges whether the eye movements fall within such AOIs (target objects) at each frame. Developers can access two kinds of gaze information as at the interaction primitive layer: participants' *target objects* and respective *gazing intervals* by just setting AOIs. In addition, our platform also provides *eye-coordinates data* on the display monitor as at the raw data layer in order for developers to interpret them by defining their own specific rules.

(3) *Writing action:* Writing action is observed in various context of learning situations such as problem-solving processes, copying down others' insightful comments, and writing up ideas advanced by participants. In order to incorporate any such writing actions by participants, our platform is designed for handling a digital pen device. The *timing of writing* information is captured and provided for developers when a participant starts touching and holds off the pen point.

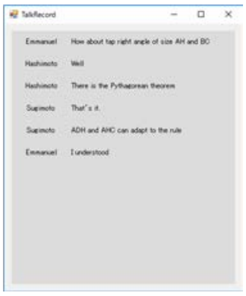
(4) *Head movement:* In addition to the gaze interaction, the instantaneous reaction by the head movement, e.g. inclining one's head and nodding, plays an important role to regulate and further the conversation (Kita and Ide, 2007). In order for developers to access such social signal, our platform provides head move information as *head direction data* (*roll, pitch, yaw*) in the head-centered coordinate system by using depth camera.

All the verbal and non-verbal information as explained above are detected on each client side in parallel and sent to the server. Developers can access this information by registering target information required for developing their learning support tools.

3.3 Message Processing

In the platform, various types of messages are exchanged among the server and clients via *message communication module*, e.g., user information for authentication processing, verbal and non-verbal information detected from devices, etc. In order to satisfy *R2*, we employ a message processing mechanism that discriminates all the messages according to their type. Figure 3 represents the class hierarchy of message classes. All the types of messages inherit *'msg_type'* property from the 'Message class' for identifying their type. In our platform, messages are classified into two categories. One is 'SystemMessage' (SY_M) used for authentication processing such as login to or logout to manage a collaborative learning session. We predefined SY_M statically. The other is 'SessionMessage' (SE_M) which developers can access in order to develop their specific learning support tools. Moreover, SE_M fall under the following three types:

- SessionInfoMessage (SI_M): are generated by the platform when a user participates in or leaves a collaborative learning session. Based on this type of messages, learning support tools can handle who participates in the session.
- MultimodalDataMessage (MD_M): correspond to verbal and non-verbal information provided by the platform as shown in Table 2. The data detected from each sensing device is provided for developers as subclass messages of MD_M such as 'StartWriting', 'GazeIn', etc. These messages are once sent to the server, and distributed to client sides in the collaborative learning session. Developers can register message types of MD_M as necessary for each learning support tool in order to receive data and implement specific multimodal interpretation processing.



Emmanuel	How about tap right angle of size AH and BC
Hashimoto	Well
Hashimoto	There is the Pythagorean theorem
Sugimoto	That's it.
Sugimoto	ADH and AHC can adapt to the rule
Emmanuel	I understood

Figure 4. Talk Record Tool.

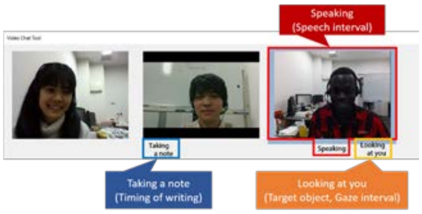
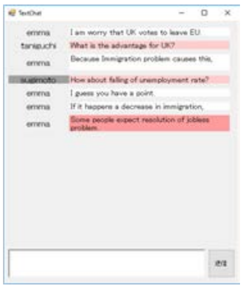


Figure 5. Non-verbal Aware Video-chat Tool.



emma	I am worry that UK votes to leave EU
taniguchi	What is the advantage for UK?
emma	Because Immigration problem causes this,
sugimoto	How about falling of unemployment rate?
emma	I guess you have a point.
emma	If it happens a decrease in immigration,
emma	Some people expect resolution of jobless problem.

Figure 6. Gaze Aware Text-chat Tool.

- LearningSupportToolMessage (LT_M): defined by developers when they implement specific learning support tools. Developers define them as subclass messages of LT_M such as 'ChatTextMessage', 'WritingCoordinateMessage', etc. As same as MD_M, developers can specify their own specific message types of LT_M as necessary for each learning support tool in order to enable specific multimodal interpretation.

To communicate all the messages between the server and clients, they are serialized into JSON format inside the *message communication module* as shown in Fig.2. Received messages are properly discriminated according to each '*msg_type*', including those defined by developers at the client side.

4. Usefulness of the Platform

In this section, we discuss the usefulness of our platform. In section 4.1, in order to demonstrate that our platform meets the requirements *R1* and *R2* and its usefulness, we give a few examples about how learning support tools utilizing verbal and/or non-verbal information provided by the platform are realized. In section 4.2, we illustrate a use case of multimodal aware learning support tools to demonstrate the usefulness of the platform as an execution environment for visualizing data captured in a CSCL session.

4.1 Usefulness from the Viewpoint of Developing Learning Support Tools

Figures 4, 5 and 6 show examples of learning support tools utilizing verbal and/or non-verbal information, implemented on the platform.

- *Talk record tool* (Fig.4): shows a sequence of speech-recognition results corresponding to a participant's utterance as described in section 3.2. We confirmed that it can be realized by just registering 'ContentOfUtterance'(*content of utterance*) type message provided by the platform as subclass message of MD_M, and implementing the function that appends its values (a couple of a participant's name and a content) to the result area when the tool deployed at each client receives the message.
- *Non-verbal aware video-chat tool* (Fig.5): the difference between non-verbal aware video-chat tool and ordinary one is that with the former (Non-verbal aware tool), both participants and non-verbal aware tools are aware of other participants' behaviors such as who is speaking, who is writing as well as who is looking at whom in real-time, whereas with the latter (ordinary tool), only participants are aware of one another without knowing who is looking at whom explicitly. In the platform, audio and video streaming data from microphones and web cameras are sent via the *stream communication module*. Thus, developers can realize an ordinary video-chat tool easily by just registering SI_M message to receive, detect, login or logout from the session, and implementing the function of displaying/hiding the participant's video object according to the SI_M. The non-verbal aware video-chat tool can be constructed by extending the basic video-chat tool. It requires for developers first to implement functions for setting an AOI region on each video object, and to register 'StartGazing' and 'StopGazing'(*gazing interval* and *target object*), 'StartSpeaking' and 'StopSpeaking'(*speech interval*) as well as 'StartWriting' and 'StopWriting'(*timing of writing*) type messages to receive provided non-verbal information as subclass messages of MD_M. Then, they can realize the tool by implementing the following

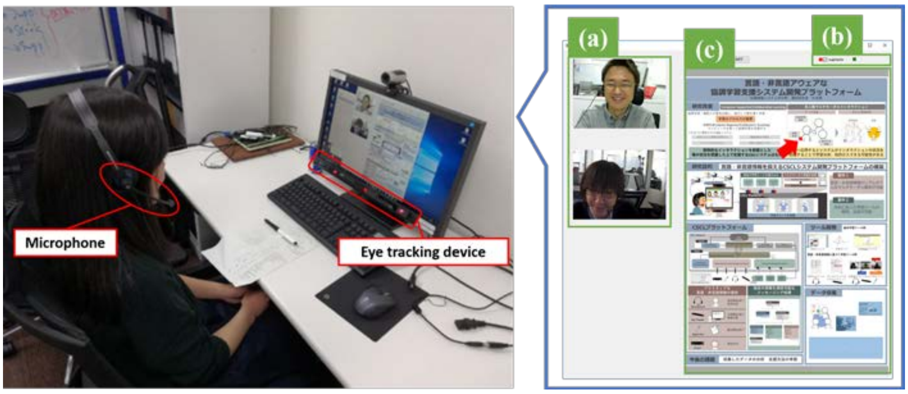


Figure 7. Shared Poster Tool Running on the Platform.

functions: highlighting the video frame of speaker and displaying labels when others are writing or looking to him/her, in accordance with each *msg_type*.

- *Gaze aware text-chat tool* (Fig.6): The tool is realized by extending basic text-chat tool so as to be aware of participants' gazing behaviors. The tool allows participants not only to exchange text messages but also to grasp text messages which participants are focused on at the moment. The basic text-chat tool can be easily built by just defining specific 'ChatTextMessage' type message as a subclass of LT_M which includes each content of the text message, and implementing functions to send and receive the messages of this message type. By extending the basic text-chat tool, we confirmed that the gaze aware text-chat tool can be realized by just doing taking the following routine: implement a function that sets an AOI region on each text-chat message object which is generated when the tool deployed at each client receives this type of message from the server; register 'StartGazing' and 'StopGazing' (*gazing interval* and *target object*) types message as subclasses of MD_M. Finally, We could realize the tool by implementing the function that highlights the background of text messages gazed at by other participants. In the case that a text message is gazed at by plural participants, its background is deeply colored to highlight according to the number of gazing participants.

By illustrating the development of learning support tools, we confirmed that the platform satisfies the two requirements: *R2* is satisfied, since SI_M provided in the platform and LT_M defined by developers properly communicate through the learning support tools. In addition, *R1* is satisfied, since the developed tools correctly work by receiving (accessing) several kinds of primitive verbal and non-verbal information provided by the platform as subclass messages of MD_M. Furthermore, we understand the usefulness of the platform for system developers, since we showed that they can realize verbal and non-verbal aware CSCL tools without having to spend time and energy in implementing low level functions for networking and sensor signal processing.

4.2 Usefulness of Multimodal Aware CSCL Platform

As a use case of multimodal aware CSCL platform in a practical collaborative learning, we implemented a shared poster tool on the platform. Figure 7 shows the situation where three participants (*A* who made the poster, *B* and *C*) take part in a collaborative learning for discussing about their collaborative research related poster contents using the developed tool. *A*, *B* and *C* can communicate with one another through video-chat (Fig.7(a)), get the control of handling the shared pointer (Fig.7(b)), and point to arbitrary locations to focus on by dragging the mouse pointer (Fig7.(c)). The poster tool also has functions that capture and record participants' gaze target objects (AOI regions on the poster) and speech intervals of their respective utterances.

Figure 8 shows set AOI regions (colored 17 regions on the poster), participants B and C as viewed by A (Fig.8(a)), speech intervals of each participant's utterances and gaze information along the timelines captured by the shared poster tool running on the platform. In the timelines (upper right of Fig.8) utterance information is represented as a series of red-colored intervals, each of which corresponds to the participant's speech interval, whereas upper gazing information of each participant is represented as multi-colored intervals, each of which corresponds to the participant's gazing interval to an AOI region on the shared poster (the color of the interval is the same as the AOI the learner gaze

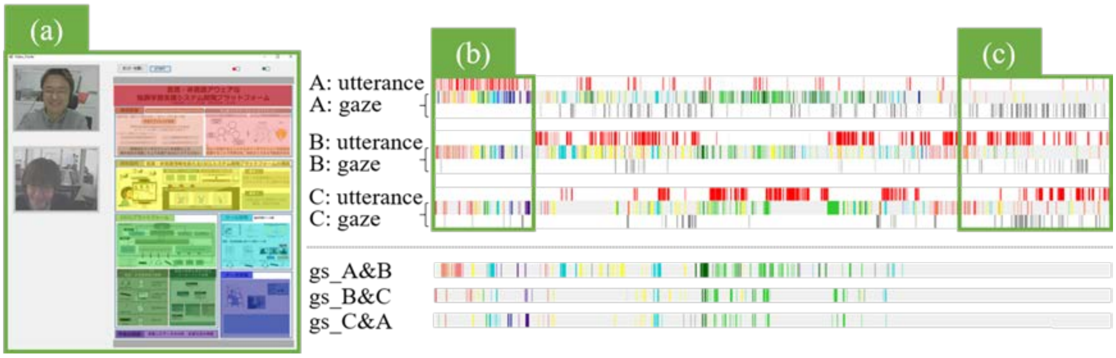


Figure 8. Data of the Collaborative Learning Conducted by Three Participants.

at) and finally gray-colored intervals represents the interval during which the participant gaze at other participants' video object.

While this session took about 30 minutes, participants could focus on their discussion without any disturbance caused by the communication control of the platform. In addition, as shown in the timelines, the platform could properly capture participants' utterance and gaze information throughout the session. Each of the timelines as in the lower right of Fig.8 represents the situations where a pair of participants (A&B, B&C and C&A) gazed at same sharing objects[1] (AOIs) in the poster simultaneously, which are captured by integrating the two participants' gaze information. As this shows, they were not always gazing at the same objects, especially at the last half of the session. According to the timelines, we could infer that the participant *A* first explained her poster contents in a step-by-step manner, while both *B* and *C* followed her explanation with gazing at AOI regions on the poster contents (Fig.8(b)). In the end of the session, participants mainly discussed not the poster contents but the future work of the research. This process appears in the timelines (Fig8.(c)); while there is no interval where plural participants were gazing at the same sharing objects on the poster, *A* was mainly observing the discussion between *B* and *C* who actively exchanged their opinions.

As described above, we confirmed that the developed learning support tool successfully run on the platform. Furthermore, we showed the usefulness and potential of the multimodal aware platform for inferring learning processes by utilizing the captured verbal and non-verbal information.

5. Discussion

5.1 Contribution as CSCL Systems Development Platform

As described in the previous sections, our platform make it easier for developers to implement learning support tools which utilize multimodal information through a simple authoring flow consisting in registering target message types which are subclasses of SI_M, MD_M, or LT_M, and adding functions that deal with received messages.

The platform also allows developers to define their unique multimodal interpretation processing by using registered messages, such as to detect situations where a participant is taking notes while an another talks by combining utterance and writing, and the situation that who are gazing at same object by referring to plural participants' gaze targets information as shown in section 4.2.

In total, the platform as a CSCL system development environment contributes to eliminating developers' workloads to develop their learning support tools, since they do not need any more to care about implementing lower level functions such as multimodal data detection, authentication processing and session management.

5.2 Contribution to the Research Field of Multiparty Multimodal Interaction

In the research field of multiparty multimodal interaction, several studies have been conducted to analyze small-group 'face-to-face' interactions based on multimodal verbal and non-verbal information such as speech, gaze and gesture (Gatica-Perez, 2009). Otsuka et al. (2008) proposed an automatic

[1] This corresponds to the concept of 'joint attention' appearing in face-to-face communication.

identification system that estimates the visual and focus of attention (i.e., *"who is looking at whom"*), in addition to speaker diarization (i.e., *"who is speaking and when"*) by using audio and visual signals in real-time. Moreover, McCowan et al. (2004) proposed an automatic analysis method of meeting actions (e.g., monologue, discussions, and presentations) based on an interaction model between participants. In order to achieve this, they extracted a number of simple audio (pitch, energy and speaking rate) and visual (head and hand blobs) data as non-verbal features automatically derived from multiple cameras and microphones, then they employed Hidden Markov Model variations to estimate meeting actions. Hillard et al. (2003) proposed a recognition model of a specific kind of interaction in meetings (agreement vs. disagreement) based on the number of utterances of each participant and positive/negative words included in respective their utterances using machine learning techniques. These studies make practical use of multimodal interaction corpora (Carletta et al., 2005 and Sumi et al, 2011), which are collections of annotated verbal and non-verbal data of multiparty interaction, scientifically analyze and model multiparty human interaction in addition to traditional methods such as participant or video observation. In order to build an interaction corpus, it firstly requires to set the environment where several measurement devices are implemented to collect the intended multimodal data, then annotating each item of data with its proper label.

Our platform makes two major contributions to the research field of multiparty multimodal interaction; The first is that it can capture various verbal and non-verbal information in collaborative learning processes in real-time by setting relatively small-scale equipment (a computer and sensing devices for each participant as shown in Fig.7), and the second is that it makes it possible to utilize captured data including several kinds of primitive verbal and non-verbal information as multimodal interaction corpora for analyzing interaction processes in a collaborative learning situation.

Of course, we need to carefully address whether we can directly apply the findings of multiparty multimodal interaction research to implement functions for analyzing interactions via CSCL systems on our platform, since there are some differences related to the interaction environment (face-to-face vs. remote), the objective of conversations and learning, and the set of tools used during the interactions. Some of the interesting and important future works would be to clarify commonalities and differences between existing findings in multiparty multimodal interaction and the ones we will get through the use of our platform.

5.3 Potential from the Viewpoint of Learning Analytics

Learning analytics (LA) and/or educational data mining (EDM) has recently been the subject of a great deal of attention. The field aims to find out patterns from the big data being accumulated in LMS, CMS and e-learning systems in order to characterize learners' behaviors and achievements, and make use of them to predict and improve educational functionalities (Peña-Ayala, 2014). El-Halees (2009), for instance, applied data mining techniques called association, classification, clustering and outlier detection rules to the collected students' data to analyze students' behavior. Mazza and Milani (2005) proposed a system that visualizes tracking data of students' behaviors on learning materials, e.g., the history of pages visited, the number of messages read and posted in discussions, to help instructors become aware of what is happening in the learning classes.

In addition to the information used so far in the LA/EDM research field such as logs of learners' learning contents and scores, our platform can also capture and utilize exhaustive verbal and non-verbal information during the collaborative learning processes as learners' primitive interaction data. Our platform has a potential to be able to contribute as a learning analytics platform focusing on communication signals level in the interaction processes of collaborative learning, which cannot be captured in traditional LMS and CMS.

6. Conclusion

In this paper, we proposed a verbal and non-verbal aware platform for developing CSCL systems. Our platform is designed to satisfy the following two core requirements: enabling a mechanism that provides verbal and non-verbal information for system developers, and a mechanism that properly communicates various types of data including unique ones defined by developers. It has functions that manage lower level processing such as authentication processing, session management and detection processing of verbal and non-verbal information, so that developers can concentrate on developing their learning support tools according to the targeted collaborative learning activities. We provided several examples of learning support tools developed on the platform utilizing verbal and/or non-verbal information. Through using them in practical collaborative learning, we confirmed the usefulness of our platform as a CSCL system development environment, and also demonstrated its potential as a basis for learning analytics in computer supported multimodal interaction.

For future works, in parallel to the extension of our platform so as to handle other kind of nonverbal information (e.g., paralanguage information), we intend to perform collaborative learning using CSCL systems implemented on our platform in order to get insights and build up findings about computer supported multimodal interpretations. We also plan to validate the effectiveness of developed learning support tools through operating the platform. It is also an interesting question to make clear how to build learning support tools embedding multimodal interpretation mechanism to capture the interaction context and to support the collaborative learning.

References

Carletta, J. et al. (2005). The AMI meeting corpus: A pre-announcement. In *International Workshop on Machine Learning for Multimodal Interaction* (pp. 28-39). Springer Berlin Heidelberg.

El-Halees, A. (2009). Mining students data to analyze e-Learning behavior: A Case Study. *Department of Computer Science, Islamic University of Gaza PO Box*, *108*.

Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, *27*(12), 1775-1787.

Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.

Hillard, D., Ostendorf, M., & Shriberg, E. (2003). Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers-Volume 2* (pp. 34-36). Association for Computational Linguistics.

Jermann, P., Soller, A., & Muehlenbrock, M. (2001). From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In *European Conference on Computer-Supported Collaborative Learning EuroCSCL-2001* (pp. 324-331).

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, *26*, 22-63.

Kita, S., & Ide, S. (2007). Nodding, aizuchi, and final particles in Japanese conversation: How conversation reflects the ideology of communication and social relationships. *Journal of Pragmatics*, *39*(7), 1242-1254.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in human behavior*, 19(3), 335-353.

Mazza, R., & Milani, C. (2005, July). Exploring usage analysis in learning systems: Gaining insights from visualisations. In *Workshop on usage analysis in learning systems at 12th international conference on artificial intelligence in education* (pp. 65-72).

McCowan, L., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., & Zhang, D. (2005). Automatic analysis of multimodal group actions in meetings. *IEEE transactions on pattern analysis and machine intelligence*, *27*(3), 305-317.

Otsuka, K., Araki, S., Ishizuka, K., Fujimoto, M., Heinrich, M., & Yamato, J. (2008). A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization. *In Proceedings of the 10th international conference on Multimodal interfaces* (pp. 257-264). ACM.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432-1462.

Red5, Red5 Media Server, http://red5.org

Strijbos, J. W., Martens, R. L., Jochems, W. M., & Broers, N. J. (2004). The effect of functional roles on group efficiency using multilevel modeling and content analysis to investigate computer-supported collaboration in small groups. *Small Group Research*, 35(2), 195-229.

Sumi, Y., Yano, M., & Nishida, T. (2010, November). Analysis environment of conversational structure with nonverbal multimodal data. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (p. 44). ACM.