



# Multimodal Analytics for Real-Time Feedback in Co-located Collaboration

Sambit Praharaj<sup>1</sup>(✉) , Maren Scheffel<sup>1</sup> , Hendrik Drachslers<sup>1,2,3</sup> ,  
and Marcus Specht<sup>1</sup> 

<sup>1</sup> Open Universiteit, Valkenburgerweg 177, 6419AT Heerlen, Netherlands

<sup>2</sup> DIPF, Schloßstr. 29, 60486 Frankfurt am Main, Germany

<sup>3</sup> Goethe Universität, Robert-Mayer-Str. 11-15, 60629 Frankfurt am Main, Germany  
{sambit.praharaj,maren.scheffel,hendrik.drachslers,marcus.specht}@ou.nl

**Abstract.** Collaboration is an important 21st century skill; it can take place in a remote or co-located setting. Co-located collaboration (CC) is a very complex process which involves subtle human interactions that can be described with multimodal indicators (MI) like gaze, speech and social skills. In this paper, we first give an overview of related work that has identified indicators during CC. Then, we look into the state-of-the-art studies on feedback during CC which also make use of MI. Finally, we describe a Wizard of Oz (WOz) study where we design a privacy-preserving research prototype with the aim to facilitate real-time collaboration in-the-wild during three co-located group PhD meetings (of 3–7 members). Here, human observers stationed in another room act as a substitute for sensors to track different speech-based cues (like speaking time and turn taking); this drives a real-time visualization dashboard on a public shared display. With this research prototype, we want to pave way for design-based research to track other multimodal indicators of CC by extending this prototype design using both humans and sensors.

**Keywords:** Collaboration · Feedback · CSCL  
Intervention · Multimodal indicators · Multimodal learning analytics

## 1 Introduction

Collaboration is an important skill in the 21st century [10]. It can take place in different settings and for different purposes: collaborative meetings [17, 36, 38], collaborative problem solving [34], collaborative project work [7, 8], collaborative programming [15] and collaborative brainstorming [37]. Some are in co-located and some in remote settings. “The requirement of successful collaboration is *complex, multimodal, subtle*, and learned over a lifetime. It involves *discourse, gesture, gaze, cognition, social skills, tacit practices*, etc.” [emphasis added] [35]. Moreover, in each context, the indicators of collaboration vary. For instance, in collaborative programming pointing to the screen, grabbing the mouse from the partner and synchrony in body posture are relevant indicators for

good collaboration [15]; whereas in collaborative meetings gaze direction, body posture, speaking time of group members are more relevant indicators for good collaboration quality [17, 36, 38]. Thus, it is essential to understand what the different types of collaboration and their purpose are and what are the relevant indicators. These indicators help to formulate the intervention or feedback mechanism to facilitate collaboration [2, 5, 30]. Moreover, engaging in a collaborative task does not essentially build collaborative skills [12]; rather on-time feedback encourages self-reflection [23]. The type of feedback is also dependent on the goal of the task which can be to evaluate collaboration as a process [2] or collaboration as an outcome (indicated by learning gain) [30] or both [30]. To understand this in-depth, we have formulated two research questions:

**RQ 1:** What *collaboration indicators* can be observed and are relevant for the *quality* of collaboration during CC?

**RQ 2:** What are the state-of-the-art *feedback* mechanisms that are used during CC?

There has been a dearth of studies on automated multimodal analysis in non-computer supported environments [40]. Considering the time and effort required to build a sensor-based automated system which can also give real-time feedback, we chose to create a WOz research prototype which can integrate human observers and existing sensor technology. This enables us to study different CC settings with a variety of multi-source multimodal indicators coming from automated sensors as well as human observers.

The remainder of the paper is structured as follows: in the related work (Sect. 2) section we answer RQ 1 and RQ 2; it is followed by an explanation of our prototype design based on the WOz study (Sect. 3); this is followed by a discussion (Sect. 4) of the answers to our research questions; finally, a conclusion (Sect. 5) is drawn and we throw some light on future work and open questions to be answered.

## 2 Related Work

In this section, we will first analyze related work according to the different indicators used during CC from multiple modalities; and secondly review the different feedback mechanisms used during CC.

### 2.1 Multimodal Indicators During Co-located Collaboration

Different categories of verbal and non-verbal indicators have been used in the literature to measure collaboration quality ranging from tangible interaction, different speech-based cues, to gaze and eye interaction. Schneider and Blikstein [27] used Tangible User Interface (TUI) for pairs of students to predict learning gains by analyzing data from multimodal learning environments. They tracked the gesture and posture using a Kinect Sensor<sup>1</sup> (Version 1) which can

---

<sup>1</sup> An integrated sensor tracking simultaneously infrared, depth, audio and video.

track the posture and gesture of a maximum of four students at a time based on their skeletal movements. They found that the *hand movements* and *posture movements* (coded as active, semi-active and passive) are correlated with learning gains. The more active a student is, the higher is the learning gain. Even the number of transitions between these three phases was a strong predictor of learning. Students who used both hands showed higher learning gains. Some of the activities that were logged by the TUI, like the frequency of opening the information box in the TUI can be correlated with learning gain. All these features were fed into a supervised machine learning framework to predict learning gain. Similarly, Martinez-Maldonado et al. [21] used TUI indicators for group work based on the log data generated and the gesture and posture of group members around the TUI.

Other works detected non-verbal cues during collaboration without a TUI. Stiefelhagen and Zhu [36] tried to detect the impact of *head orientation* on the gaze direction in a group round table meeting with four members. They found that on an average 68.9 % of the time head orientation can estimate gaze direction. Moreover, attention focus of group members can be easily predicted 88.7 % of the time using head orientation as the only input. Similarly, Cukurova et al. [7] performed an experiment on 18 members in six groups of three members each to detect non-verbal cues of collaboration using human observation. *Hand position* (HP) and *head direction* (HD) was a good predictor of competencies in Collaborative Problem Solving (CPS). They extended this work and formed the NISPI framework [8] using HP and HD as non-verbal indicators. These indicators were obtained during a prototype design by students (11–20 years old) using the Arduino toolkit. Then, they were coded for each student as: 2 (*active*) if a student is interacting with the object for problem solving, 1 (*semi-active*) if the head of the student is directed towards an active peer and 0 (*passive*) for all other situations. Using this coding, different collaboration dimensions like *synchrony*, *individual accountability* (IA), *equality* and *intra-individual variability* (IIV) were formed. High competencies of CPS was detected if high levels of synchrony, IA and equality is detected in the groups.

Speech-based cues are an integral part of any collaborative task. Lubold and Pon-Barry [19] found that *proximity*, *convergence* and *synchrony* are different types of coordination cues obtained from the speech features (like intensity, pitch and jitter) of the pair of students collaborating. It helped them to detect rapport between group members. It was observed from correlation analysis that proximity, convergence and synchrony measured using pitch can be a good predictor of rapport between the group members during collaboration. Students also self-reported rapport which was compared and collaboration levels were determined. Bassiou et al. [4] assessed collaboration among students solving math problems automatically. They used *non-lexical speech* features; thereby, preserving the privacy. They used a combination of manual annotation and Support Vector Machine (SVM) to predict the collaboration quality of the group. Types of collaboration marked are: Good (all 3 members are working together and contributing to the discussion), Cold (only two members are working together),

Follow (one leader is not integrating the whole group) and Not (everyone is working independently). This coding was based on two types of engagement: simple (talking and paying attention) and intellectual (actively engaged in the conversation). They found that the combination of the speech-activity features (i.e., *solo duration*, *overlap duration of two persons*, *overlap duration of all three persons*) and speaker-based features (i.e., *spectral*, *temporal*, *prosodic* and *tonal* features of speech) are good predictors of collaboration. Simple indicators like the speaking time of each member can also be a good indicator of collaboration [2, 5]. Even a mixture of verbal and non-verbal indicators along-with *physiological signals* like skin temperature [24] can be a good collaboration indicator [18, 20].

Besides, *eye gaze* can be an indicator of collaboration quality. Some researchers [16, 25, 28] while using eye gaze analysis found that (JVA) *Joint Visual Attention* (i.e., the proportion of times gazes of individuals are aligned by focusing on the same area in the shared object or screen) is a good predictor of the quality of collaboration of a group which is reflected by the groups performance. Moreover, Schneider and Pea [28] showed that JVA can be used as a reflection mechanism in remote settings to show each student their partners gaze patterns in real-time to improve collaboration. Schneider et al. [30] got the same results by replicating the experiment in a co-located setting. The work by Schneider and Pea [29] used JVA, network analysis and machine learning to determine different dimensions of a good collaboration like *mutual understanding*, *dialogue management*, *division of task*, *signs of coordination* as outlined by Meier et al. [22].

Moving on to the different purposes in which collaboration has been studied, Spikol et al. [33, 34] studied collaborative learning specifically in the context of Collaborative Problem Solving (CPS). They tracked the distance between *hand movements* and *faces* of group members. Later the recorded video streams were coded by experts with 0 (for passive), 1 (for semi-active) and 2 (for active) based on different combinations of head and hand positions for training the machine learning classifier for predicting the quality of collaboration. Recent work by Chikersal et al. [6] dives deep into the deep structure of collaboration in dyads. They found that synchrony in facial expressions correlated with collective intelligence of the group but not significantly correlated with the synchrony of electrodermal activity of members. Another work by Grover et al. [15] studied CPS in a pair programming context based on a pilot study. They captured data from different modalities (i.e., video, audio, clickstream and screen capture) unobtrusively using Kinect. For initial training of the classifiers using machine learning, experts coded the video recordings with three annotations (i.e., High, Medium and Low) when they found evidences of collaboration between the dyads. These evidences include *pointing to the screen*, *grabbing the mouse* from the partner and *synchrony in body position*. Later this classifier could predict the level of collaboration.

Moreover, post-hoc coding with the help of human coders has been an effective method followed for a long time to detect different indicators of collaboration. Davidsen and Ryberg [9] videotaped the work of pairs making a collaborative

**Table 1.** Overview of studies on co-located collaboration.

References	Indicators	Goal
[27]	Hand movements, posture & TUI logs	Post-hoc analysis of indicators on learning
[30]	Joint Visual Attention (JVA)	JVA indicates learning
[34]	Distance between hands & faces	Extraction of multimodal features during collaboration
[15]	Pointing, body position & grabbing mouse	Post-hoc classification of collaboration
[2]	Total speaking time	LED display to regulate audio participation in real-time
[37]	Number of ideas	Real-time metaphorical feedback to support CB
[5]	Total speaking time	Conversation clock will regulate the equity of conversation in real-time
[8]	Hand position and head direction	Build a non-verbal indicator framework for collaboration
[19]	Intensity & pitch of sound, self reports	Detect collaboration levels based on rapport obtained from audio cues & self-reports
[4]	Speech overlap duration, no overlap duration, spectral, temporal, prosodic & tonal speech features	Predict collaboration quality from audio cues
[9]	Dialogue, gesture, posture & gaze	Detect indicators of collaboration from videotaped recordings of collaboration tasks
[26]	Eye contact, posture & amplitude of voice	Detect indicators of collaboration from videotaped recordings of collaboration tasks

discussion around a touch screen measuring “The size of one meter”. The pair was trying to translate the design from graph paper to the touch screen to measure one meter. They found that *body movements*, *language* and *gestures* can be helpful to discover different facets of collaboration. Similarly, Scherr and Hammer [26] observed videotaped groups and identified four clusters based on the collaborative behaviour from both verbal and non-verbal indicators (like *eye contact with peers*, *straight posture*, *clear and loud voice*, etc.). Besides, some works [32, 37] considered *epistemological* aspects of collaboration during brainstorming where the number of ideas generated by each member was the indicator of quality of collaboration. Detecting individual attention levels in classroom from the responses to questions (i.e. epistemological) is also common [39].

In summary, collaboration indicators can vary from non-verbal, verbal, physiological to log files obtained from shared objects like TUI or computers. It depends on the context. Table 1 shows the overview of the multimodal indicators detected. We can find two types of co-located collaboration indicators, i.e., *social* (verbal, non-verbal and physiological) and *epistemological* (logs, ideas).

## 2.2 Feedback During Co-located Collaboration

Using these multimodal indicators, different feedback mechanisms have been developed in the past to facilitate CC. Kulyk et al. [18] designed a mechanism to give real-time feedback to participants in group meetings (with 4 members) by analyzing their speaking time and gaze behaviour. The feedback was in the form of different coloured circles representing attention from other speakers measured by eye gaze, speaking time and attention from listeners. This feedback was projected on the table in-front of where each participant was sitting using a top-down projector. They performed both quantitative and qualitative evaluation to evaluate the effect of the feedback: the feedback was accepted as a positive measure by most group members; use of feedback had a positive impact on the behaviour of group members as they had a balanced participation and improved eye gaze. Terken and Strum [38] used a similar setting and feedback mechanism; they discovered that the feedback on speech increased the equity of participation in the group. But, surprisingly feedback on gaze behaviour had little effect on the interaction pattern of group members. Similarly, Madan et al. [20] used sensors to capture nodding, speech features and galvanic skin response of dyads and built a real-time group interest index. This group interest index helped them to drive a real-time feedback. This feedback showed some group characteristics in different modes: individual PDA feedback, personal audio feedback, haptic feedback in the shoulder and public shared projected display. They studied these group characteristics in different contexts like speed dating and brainstorming sessions.

Some simpler versions of feedback which leverage the audio cues (like speaking time) during collaboration have proved effective in the past. For instance, Bachour et al. [2] performed an experiment to measure audio participation where each group (with 3–4 members) performed a task around a smart table. It gave them real-time feedback during the task by glowing different coloured LED lights for each member. The number of LED lights that glowed for each colour denoted the total speaking time for that member. They found that a real-time feedback helped to maintain the equity of audio participation among the members. Another similar approach was used by Bergstrom and Karahalios [5] with the help of a conversation clock. In this clock, different coloured concentric rings represented spoken participation of each member in the 4 member group. The bars and the dots in the ring denoted the length of conversation and periods of silence respectively.

Moving on to the epistemological aspect of collaboration, Tausch et al. [37] used an intuitive metaphorical feedback moderated by human observers during *collaborative brainstorming*. Three members in each group performed the task. The group members were supposed to discuss a certain topic and their collaboration was measured by the number of ideas generated. A comparison metric for collaboration such as a baseline was calculated as the average number of ideas generated by all members. Using this baseline, each group member was marked as below average or above average depending on the number of ideas generated by each member. Then the human observers controlled the public shared display

which showed a *metaphorical garden*. Each group member was represented by a flower and the group was displayed as a tree with leaves, flower and fruit. The growth of the flower and the tree symbolized the participation (measured by the contribution of ideas) of the individual and the group respectively. More balanced participation was shown by a well grown tree with leaves, fruits and flower. If a group was having unbalanced participation for a long time then lightning flashes were shown in the group garden. Another example of feedback during collaborative brainstorming was implemented by Shih et al. [32]. It supports collaborative conceptual mapping to discuss a topic and organize the ideas.

Besides the use of visual and haptic feedback was effective in some collaboration tasks around a TUI. Anastasiou and Ras [1] gave real-time textual and haptic feedback to each group consisting of 3 members working around a TUI. The group members were needed to use different objects and find the desired power consumption using the TUI. At the end, they used a questionnaire and found that most participants of the experiment favoured the use of both visual and haptic feedback over audio feedback. Martinez-Maldonado et al. [21] used a TUI and gave real-time feedback on group performance for the teachers in tablets so that they can intervene when needed and can also make a post-hoc reflection after the task is over.

Use of external sensing devices to facilitate collaboration during meetings has proved its worth before. Kim et al. [17] used a sociometric badge<sup>2</sup> which acted as a meeting mediator to capture audio and postures during meetings of 4 members in one group. This badge bridged the gap of dominance and increased the equity of participation among the group members using a real-time feedback on their personal mobile phones. This feedback showed a circle in the middle of a screen connected by four lines to small squares in each corner of the screen representing the individual group members. The colour and position of the circle denoted the interactivity of the group. When the group had a balanced participation then the circle was darker in colour and in the centre of the screen. The thickness of lines connecting the circle represented the speaking time of each group member. Apart from the personal mobile display to give feedback, Balaam et al. [3] used an ambient display showing a coloured circle visualization based on the non-verbal indicator of synchrony during a collaborative task of calendar planning. DiMicco et al. [13] used a shared group display to influence the speaking participation of each group member during a group activity.

In summary, most of these studies were in controlled conditions with small groups consisting of dyads and triads only. Table 2 shows the overview of feedback mechanisms used during co-located collaboration. Some real-time feedback mechanisms acted as a mere reflection for the group to self-regulate instead of an actionable feedback; while others used a post-hoc analysis for the teachers (or facilitators) to reflect on the group activity. The mode of display varied from a public display to smart phone display.

---

<sup>2</sup> An electronic sensing device worn around the neck that can collect and analyze social dynamics.

**Table 2.** Overview of studies on co-located collaboration feedback.

References	Indicators	Feedback
[17]	Total speaking time & body posture	Graphical with coloured shape and lines using personal mobile screens
[37]	Number of ideas	Metaphorical as a groupgarden using public shared display
[18]	Speaking time & eye gaze	Graphical with coloured concentric circles using public table-top private projection
[38]	Speaking time & eye gaze	Graphical with coloured concentric circles using public table-top private projection
[20]	Nodding, speech features & galvanic skin response	Graphical group characteristics using audio, haptic, PDA and public shared display
[2]	Total speaking time	Coloured LED light using public shared table top LED display
[5]	Total speaking time	Coloured concentric rings with lines and dots using public shared table top display
[3]	Pointing	Coloured circle visualization using ambient display
[21]	Log data about different actions performed with the TUI	Pie chart and other statistical charts using private tablet for teachers
[1]	Log data about content knowledge from TUI	Textual and haptic using public TUI display
[13]	Total speaking time	Coloured bar charts using public shared display

In a nutshell, most of the studies in *related work* are in controlled conditions and using specialized furniture, TUI and badges. These settings can be suitable for adhoc CC which can be difficult to adapt in a dynamic setting. They also do not cater to the privacy and fairness of individuals. Most of these studies employ human observers as post-hoc annotators for coding videos to detect traces of collaboration. To tackle these issues, we devise a human-based prototype where privacy, in-the-wild setting and dynamic design is at the centre of our WOz study.

### 3 A WOz Study: Designing the Research Prototype

Based on our analysis, we aimed for creating a flexible research infrastructure that allows us to study feedback in CC making use of different indicators and combining them in different feedback instruments and media. We followed a design-based approach focusing on a specific type of meeting and evaluated different types of indicators, human-observer interfaces, as well as feedback mechanisms. The main components of our research prototype are a defined set of

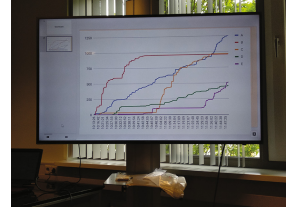




**Fig. 1.** Meeting room



**Fig. 2.** Annotator room



**Fig. 3.** Public display

indicators and sensors, a user interface for CC observation managed by human observers, as well as a set of feedback components.

### 3.1 Experimental Context

We performed the experiments during three PhD meetings with 3–7 members in each meeting in the room as shown in Fig. 1. Due to the frequent availability of these meetings and ease of not designing the task per se, we chose them. Our main focus was to execute the study *in-the-wild* and preserving the *privacy*. Thus, we used a human annotator who was present in the adjacent room separated by a one-sided transparent wall as shown in Fig. 2. Although it is difficult to see in the picture, the visibility through the wall from the side of the annotator was transparent; while the visibility from the meeting room was opaque. A microphone was used to listen to the conversation in the other room but audio was not recorded. The real-time feedback was shown on a big shared public display in the meeting room (as depicted in Fig. 3) which was managed by the annotator. The real-time feedback visualization could make use of observation data from the human observer and also visualize raw-data, e.g. the audio volume of the group work. The collaborators got a virtual sense of being tracked by a microphone automatically when they saw the changing real-time feedback of their speaking participation on the screen.

### 3.2 Data Logging

For the sake of clarity in data logging, we have segregated the multimodal channel annotation into verbal and non-verbal (i.e., gestures and postures) channels and identified different non-verbal indicators as: looking at laptop or peers; looking down; looking at the feedback; typing with laptop; and making different hand gestures. The verbal indicators are: occurrence, pauses, overlaps, interruptions in speech; affirmatives in speech; and asking questions. But, to ease the logging process for the human annotator, we chose to only focus on the simpler observable audio cues which is the speaking time and turn taking of each group member in a first study. The speech-based cues are ubiquitous in any collaboration and non-verbal cues may be difficult to monitor for one annotator in a large group setting. The annotator was seeing the annotation interface embedded in a

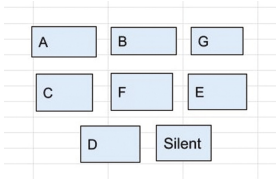


Fig. 4. Annotation interface

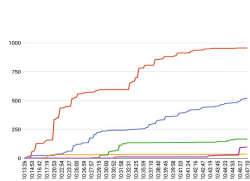


Fig. 5. Mid feedback

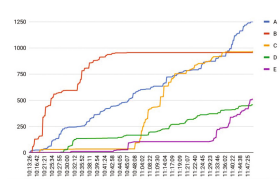


Fig. 6. End feedback

Google sheet as shown in Fig. 4. To preserve the *privacy*, we gave the annotator a coding sheet where each collaborating member was given an alias name from the English alphabet. Moreover, each participating member signed a consent form. Whenever a person starts speaking, the annotator pressed the corresponding button in the interface which automatically creates a cell in the Google sheet with the start time and name of that person. Whenever the annotator presses another person’s button, the end time of the previous person is registered in the sheet. This was possible as the buttons were coupled with a JavaScript to perform the operation. To ensure the reliability of the coding scheme, we had a provision to include multiple annotators but did not use it for our experiments as it involved only simple clicking of a button.

### 3.3 Modeling Participation During Collaboration

The sheet interface was connected to a chart embedded in Google Slides which was updated in real-time when a value is entered by pressing a button. The other columns in Google sheet were automatically populated based on the defined formula which calculates the cumulative speaking time of each member from the beginning of the meeting. Figure 5 shows the group dynamics after the first 30 min during a meeting using a line chart as displayed during the meeting on the big public shared display in the room. The times shown on the horizontal axis is the plot time obtained from the end time of speaking of a member. The value in vertical axis is the total speaking time in seconds from the beginning of the meeting. Figure 6 shows the status of the line chart at the end of the meeting. Here, the speaking time and turn taking represented the participation of each group member. We also collected oral feedback from both the annotators and the collaborators during the iterative design phase.

### 3.4 Results

From our first three iterations in the PhD meetings, we developed a first prototype for analyzing turn-taking and speaking time feedback. Our results showed that we need a higher level annotation interface. Thus, we supported human observers in that they only need to press a button when a new person starts talking. For the visualization on the public shared display, we experimented with different visualizations of the speaking time. Based on participants’ feedback we

altered the display format from an original pie chart to a line chart for displaying the development of the conversation over time. An example of the feedback at different times of a meeting can be seen in Figs. 5 and 6. We can observe that speaker B, who is a second year PhD student, dominates the conversation in the first 30 min; it was his turn to speak regarding his PhD project at that time. But, from that time on-wards he stops to participate in the meeting; indicated by the line parallel to horizontal axis in Fig. 6. We can also observe at the end of the meeting that speaker A, who is the promotor, has spoken the most and changed turns very often to intervene during the meeting; the turn-taking was evident from the frequent change of the shape of the line indicated by small or large spikes.

## 4 Discussion

**RQ1: On the multimodal indicators during CC indicating collaboration quality** — Based on the literature study, we discovered different multimodal indicators during CC in multiple contexts. They can be grouped into *social* (i.e., verbal, non-verbal and physiological) and *epistemological* (i.e., ideas and data logs) indicators. For detecting the social indicators, sensors have been used in past works. But, for detecting the epistemological indicators human help was required as it is difficult for sensors to automatically detect the number of ideas generated from speech by understanding the semantics.

**RQ2: On the feedback during CC** — Feedback during CC is either real-time (for reflection or guiding) or post-hoc (for the purpose of reflection). This brings into the picture two *stakeholders*: the teachers (or facilitators) and the group members. We need this distinction as it will help in designing the feedback. Some works used TUI and other electronic mediums like Interactive White Boards (IWB) and tablets during collaboration which requires a lot of preparation before a collaborative task. Therefore, it is difficult to use it in real-world dynamic settings. Besides, there is a trade-off between personalization for the group and privacy. More personalized feedback meant for the whole group is less privacy preserving. Thus, there should be a decision on the level (i.e., group, individual or both) of feedback to be shown depending on the circumstances at hand.

**On the research prototype to give real-time feedback** — We take a step in building an initial prototype design with the aim to facilitate real-time collaboration during meetings. We were successful in building a click-based interface for the annotator which also reduces memory overhead. This helps us to create a hybrid setup without building an actual automated sensor-based system to experiment with different types of real-time feedback mechanisms during CC. We can later use these insights to build the sensor-based or hybrid setup. Here, we can build individual components in a modular fashion to track other indicators of collaboration quality; and integrate them to a single dashboard.

## 5 Conclusions and Future Work

Collaboration being an important skill and ubiquitously present in our day to day activities, we try to look into the different collaboration indicators in various contexts in the literature. We find different types of indicators like gaze, speaking time, posture, gesture, number of ideas generated, etc. Then we look into the impact of feedback during collaboration and find that visual real-time feedback has some impact on the collaboration like improving the equity of audio participation. This feedback can range from private displays (like PDA, mobile phones) to a more public one (like TUI, shared display).

Based on this overview, we took a step further and built a real-time feedback prototype during collaboration based on a privacy-preserving WOz study in-the-wild. Here, we study collaboration during co-located PhD meetings using human observers acting as a proxy for sensors. We find that the human observers could easily track ‘who spoke when and for how much time’ by pressing a button.

As future work suggestions, we need to define the *goal* and *outcome* of the collaboration task and make it clear in the evaluation criteria as to whether we measure collaboration as a process, outcome or both. Then, we can focus on the feedback mechanisms for facilitating collaboration. We can also borrow some insights from the mapping of multimodal data to feedback in an individual learning context [11]. The feedback can be: human based, sensor based or a hybrid of both. We need to decide the type (number of pointing gestures, speaking time, number of interruptions, number of eye contact with peers, etc.), modelling (i.e., individual, group or both) and display of feedback (i.e., personal, public or both) based on *action-based research* [14] where we need to take the preliminary feedback of different stakeholders like teachers (or facilitators) and the group members. Our long term goal is to do action-based research and build a sensor-based automated (or hybrid) feedback system during CC using the currently built research prototype. Here, we can include different feedback components to identify multiple indicators of collaboration and proceed towards an automated system using deep neural networks to integrate data from multiple sensors [31].

## References

1. Anastasiou, D., Ras, E.: A questionnaire-based case study on feedback by a tangible interface. In: Proceedings of the 2017 ACM WS on Intelligent Interfaces for Ubiquitous and Smart Learning, pp. 39–42. ACM (2017)
2. Bachour, K., Kaplan, F., Dillenbourg, P.: An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Trans. Learn. Technol.* **3**(3), 203–213 (2010)
3. Balaam, M., Fitzpatrick, G., Good, J., Harris, E.: Enhancing interactional synchrony with an ambient display. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 867–876. ACM (2011)

4. Bassiou, N., Tsiartas, A., Smith, J., Bratt, H., Richey, C., Shriberg, E., D'Angelo, C., Alozie, N.: Privacy-preserving speech analytics for automatic assessment of student collaboration. In: INTERSPEECH, pp. 888–892 (2016)
5. Bergstrom, T., Karahalios, K.: Conversation clock: visualizing audio patterns in co-located groups. In: 40th Annual Hawaii International Conference on System Sciences, p. 78. IEEE (2007)
6. Chikersal, P., Tomprou, M., Kim, Y.J., Woolley, A.W., Dabbish, L.: Deep structures of collaboration: physiological correlates of collective intelligence and group satisfaction. In: CSCW, pp. 873–888 (2017)
7. Cukurova, M., Luckin, R., Mavrikis, M., Millán, E.: Machine and human observable differences in groups' collaborative problem-solving behaviours. In: Lavoué, É., Drachler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) EC-TEL 2017. LNCS, vol. 10474, pp. 17–29. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-66610-5\\_2](https://doi.org/10.1007/978-3-319-66610-5_2)
8. Cukurova, M., Luckin, R., Millán, E., Mavrikis, M.: The nispi framework: analysing collaborative problem-solving from students' physical interactions. *Comput. Educ.* **116**, 93–109 (2018)
9. Davidsen, J., Ryberg, T.: This is the size of one meter: childrens bodily-material collaboration. *Int. J. CSCL* **12**(1), 65–90 (2017)
10. Dede, C.: Comparing frameworks for 21st century skills. In: 21st Century Skills: Rethinking How Students Learn, vol. 20, pp. 51–76 (2010)
11. Di Mitri, D., Schneider, J., Drachler, H., Specht, M.: From signals to knowledge. A conceptual model for multimodal learning analytics. *J. Comput. Assist. Learn.* **34**(4), 338–349 (2018)
12. Dillenbourg, P.: What do you mean by collaborative learning? (1999)
13. DiMicco, J.M., Pandolfo, A., Bender, W.: Influencing group participation with a shared display. In: Proceedings of the 2004 ACM Conference on CSCW, pp. 614–623. ACM (2004)
14. Dyckhoff, A.L.: Action research and learning analytics in higher education. *eled* **10**(1) (2014)
15. Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., Divakaran, A.: Multimodal analytics to study collaborative problem solving in pair programming. In: Proceedings of the 6th International Conference on LAK, pp. 516–517. ACM (2016)
16. Jermann, P., Mullins, D., Nüssli, M.A., Dillenbourg, P.: Collaborative gaze footprints: correlates of interaction quality. In: CSCL 2011 Conference Proceedings, vol. 1, pp. 184–191. International Society of the Learning Sciences (2011)
17. Kim, T., Chang, A., Holland, L., Pentland, A.S.: Meeting mediator: enhancing group collaboration using sociometric feedback. In: Proceedings of the 2008 ACM Conference on CSCW, pp. 457–466. ACM (2008)
18. Kulyk, O., Wang, J., Terken, J.: Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 150–161. Springer, Heidelberg (2006). [https://doi.org/10.1007/11677482\\_13](https://doi.org/10.1007/11677482_13)
19. Lubold, N., Pon-Barry, H.: Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In: Proceedings of the 2014 ACM WS on Multimodal Learning Analytics Workshop and Grand Challenge, pp. 5–12. ACM (2014)
20. Madan, A., Caneel, R., Pentland, A.S.: Groupmedia: distributed multi-modal interfaces. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 309–316. ACM (2004)

21. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: Mfeedback: providing notifications to enhance teacher awareness of small group work in the classroom. *IEEE Trans. Learn. Technol.* **8**(2), 187–200 (2015)
22. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. *Int. J. CSCL* **2**(1), 63–86 (2007)
23. O'Donnell, A.M.: *The role of peers and group learning* (2006)
24. Pijera-Daz, H.J., Drachsler, H., Kirschner, P.A., Järvelä, S.: Profiling sympathetic arousal in a physics course: how active are students? *J. Comput. Assist. Learn.* **34**(4), 397–408 (2018)
25. Richardson, D.C., Dale, R.: Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cogn. Sci.* **29**(6), 1045–1060 (2005)
26. Scherr, R.E., Hammer, D.: Student behavior and epistemological framing: examples from collaborative active-learning activities in physics. *Cogn. Instr.* **27**(2), 147–174 (2009)
27. Schneider, B., Blikstein, P.: Unraveling students interaction around a tangible interface using multimodal learning analytics. *J. Educ. Data Min.* **7**(3), 89–116 (2015)
28. Schneider, B., Pea, R.: Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *Int. J. CSCL* **8**(4), 375–397 (2013)
29. Schneider, B., Pea, R.: Toward collaboration sensing. *Int. J. CSCL* **9**(4), 371–395 (2014)
30. Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., Pea, R.D.: 3d tangibles facilitate joint visual attention in dyads. *International Society of the Learning Sciences, Inc. [ISLS]* (2015)
31. Schneider, J., Di Mitri, D., Limbu, B., Drachsler, H.: Multimodal learning hub: a tool for capturing customizable multimodal learning experiences. In: Drachsler, H., et al. (eds.) *EC-TEL 2018. LNCS*, vol. 11082, pp. 45–58. Springer, AG (2018)
32. Shih, P.C., Nguyen, D.H., Hirano, S.H., Redmiles, D.F., Hayes, G.R.: Groupmind: supporting idea generation through a collaborative mind-mapping tool. In: *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, pp. 139–148. ACM (2009)
33. Spikol, D., Ruffaldi, E., Dabisias, G., Cukurova, M.: Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *J. Comput. Assist. Learn.* **34**(4), 366–377 (2018)
34. Spikol, D., Ruffaldi, E., Landolfi, L., Cukurova, M.: Estimation of success in collaborative learning based on multimodal learning analytics features. In: *Proceedings of the 17th ICALT*, pp. 269–273. IEEE (2017)
35. Stahl, G., Law, N., Hesse, F.: Reigniting CSCL flash themes. *Int. J. CSCL* **8**(4), 369–374 (2013)
36. Stiefelhagen, R., Zhu, J.: Head orientation and gaze direction in meetings. In: *CHI 2002 Extended Abstracts on Human Factors in Computing Systems*, pp. 858–859. ACM (2002)
37. Tausch, S., Hausen, D., Kosan, I., Raltchev, A., Hussmann, H.: Groupgarden: supporting brainstorming through a metaphorical group mirror on table or wall. In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction*, pp. 541–550. ACM (2014)
38. Terken, J., Sturm, J.: Multimodal support for social dynamics in co-located meetings. *Pers. Ubiquitous Comput.* **14**(8), 703–714 (2010)

39. Triglianios, V., Praharaaj, S., Pautasso, C., Bozzon, A., Hauff, C.: Measuring student behaviour dynamics in a large interactive classroom setting. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, pp. 212–220. ACM (2017)
40. Worsley, M., Blikstein, P.: Leveraging multimodal learning analytics to differentiate student learning strategies. In: Proceedings of the 5th International Conference on LAK, pp. 360–367. ACM (2015)