Constructing a Multi-Party Remote Discussion Corpus and Analyzing Facilitator Characteristics Using LLMs

Shoki Hatano, Kazutaka Shimada
Kyushu Institute of Technology
Fukuoka, Japan
hatano.shoki801@mail.kyutech.jp, shimada@ai.kyutech.ac.jp

Abstract

This paper presents two contributions to corpus-based research on remote communication. First, we develop the Kyutech Remote Corpus, an openly accessible Japanese corpus of multi-party remote discussions. Second, we examine differences in facilitator characteristics between face-to-face and remote discussions using a large language model (LLM). To analyze facilitator behavior, we compare the Kyutech Remote Corpus with an existing face-to-face discussion corpus. We identify facilitators using participant questionnaires and use GPT-4o to identify facilitators from discussion transcripts and explain its predictions. The generated explanations capture shared and modality-specific facilitator characteristics. In both discussion formats, facilitators frequently clarified discussion topics and organized participants' opinions. In remote discussions, facilitators more frequently managed discussion flow and promoted consensus building. In contrast, facilitators in face-to-face discussions more often deepened topic exploration and encouraged participation by other discussants. These findings suggest that facilitator behavior is shaped by interaction modality. These findings also provide empirical evidence for the design of discussion-support systems tailored to remote communication environments.

Keywords: Corpus, Multi-Party remote discussion, Facilitator

1. Introduction

Decision-making discussions play an important role in business, education, and many collaborative activities. Participants exchange ideas and reach a consensus through discussion. An effective facilitator organizes opinions, encourages balanced participation, and guides the discussion toward consensus. However, a skilled facilitator is not always present in discussions. Therefore, many studies have proposed facilitation support systems (Kirikihira and Shimada, 2018) (Nishimura et al., 2023).

Developing facilitation support systems requires conducting discussions for system verification. However, assigning a facilitator to every discussion is labor-intensive. Therefore, corpora consisting of multi-party discussions for the support system are needed. In addition, most existing facilitation support systems have focused on face-to-face discussions. Meanwhile, remote discussions conducted through video conferencing systems have become increasingly common in recent years. Previous studies have reported that participant behavior differs between face-to-face and remote discussions (Onishi et al., 2022) (Baiat et al., 2026). Therefore, remote discussion corpora are also necessary for developing facilitation support systems for remote discussions. Meanwhile, existing remote dialogue corpora mainly contain one-to-one conversations (Inaba et al., 2022) (Okahisa et al., 2022). Accordingly, constructing a multi-party remote discussion corpus is important.

Since participant behavior differs between face-to-face and remote discussions, facilitator behavior is also likely to differ. To construct facilitation support systems, it is important to understand facilitator characteristics. Understanding the characteristics in different discussion environments contributes to the development of systems that can adapt to each environment. Therefore, analyzing facilitators' behavior in face-to-face and remote discussions is significant. The constructed corpus also enables detailed analyses of facilitator behavior in discussions.

For facilitator analysis, previous studies have used classification models to identify facilitator characteristics (Shiota et al., 2018). The study identified several facilitator characteristics using manually designed linguistic features. However, the analyses depended on the predefined behavioral features. To address this gap, we use a Large Language Model (LLM) to analyze facilitator behavior. LLMs have recently demonstrated strong performance in various language analysis tasks. Unlike conventional feature-based approaches, LLMs can analyze discussion transcripts from multiple perspectives without manually designed behavioral features. Therefore, LLMs are suitable for exploring facilitator characteristics from discussion transcripts.

In this study, we construct the Kyutech Remote Corpus, a Japanese corpus of multi-party remote decision-making discussions. We then use the corpus to investigate facilitator characteristics through LLM-based analysis. We also compare facilitator behavior between face-to-face and remote discussions under the same discussion task. This comparison clarifies differences in facilitator behavior across discussion modalities.

The main contributions of this paper are as follows.

- We construct the Kyutech Remote Corpus, a Japanese corpus of multi-party remote decision-making discussions.
- We release the constructed corpus publicly on the Web[1].
- We use the proposed corpus to investigate facilitator characteristics through LLM-based analysis and clar-

[1]The Kyutech Remote Corpus: `http://www.pluto.ai.kyutech.ac.jp/~shimada/resources.html#kyutechRemote`.

ify differences in facilitator behavior between face-to-face and remote discussions.

2. Related Work

2.1. Discussion Corpus

Many studies have constructed discussion corpora to analyze human communication. Early studies mainly focused on face-to-face meetings. For example, the AMI Meeting Corpus (Kraaij et al., 2005) and the ICSI Meeting Corpus (Janin et al., 2003) have been widely used for meeting analysis research. These corpora have been widely used in meeting analysis research.

In Japan, the Kyutech Corpus (Yamamura et al., 2016) contains face-to-face decision-making discussions. This corpus consists of face-to-face discussions. However, the discussion tasks are not specifically designed for a particular dialogue format. Therefore, in this study, we construct a corpus of multi-party remote discussions based on the construction approach of the Kyutech Corpus.

Recently, researchers have also constructed corpora for remote communication (Inaba et al., 2022) (Okahisa et al., 2022). Many existing remote corpora focus on one-to-one conversations. Unlike these corpora, we construct a remote corpus focused on multi-party discussions.

2.2. Discussion Analysis

Some studies have investigated discussion participants' behavior (Huang and Nishida, 2020) (Shiota et al., 2020). They identified participant characteristics using manually designed linguistic and movement features. These studies clarified several characteristics of discussion participants in face-to-face discussions. However, remote discussions involve different recording conditions across speakers. The reason is that participants use different cameras, microphones, and other devices. As a result, nonverbal features, such as body movements and prosodic features, may reflect speaker-specific recording conditions rather than actual participant behavior. It could become an impediment to analyzing facilitator behavior. Therefore, this study analyzes facilitator characteristics by using only spoken utterances, which are less dependent on individual recording conditions.

3. Kyutech Remote Corpus

To analyze facilitator behavior in remote discussions, we constructed the Kyutech Remote Corpus. This is a Japanese multi-party remote discussion corpus. The Kyutech Remote Corpus targets discussions conducted through a video conferencing system. We designed the corpus for facilitator analysis and for the development of facilitation support systems. This section describes the data collection procedure, discussion task, annotation process, and corpus statistics.

3.1. Data Collection

The remote discussions were conducted through Zoom[2]. We recorded each discussion by using the recording function of Zoom. One participant used a PC prepared by the authors and connected to the recording equipment. The other participants used their own devices.

[2] https://www.zoom.com/ja

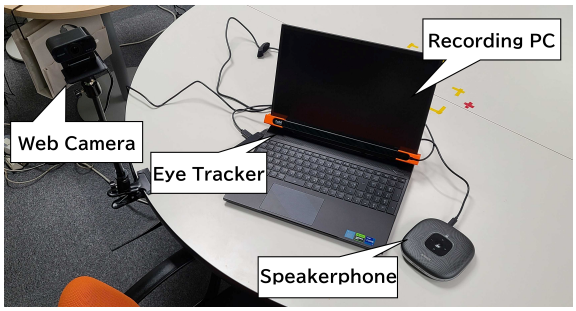

Figure 1: Recording equipment setup.

Figure 2: An example of the remote discussion.

Fig. 1 shows the recording equipment. We used a web camera (UCAM-CX80FBBK[3]) and an eye tracker (Tobii Eye Tracker 4C[4]) to record the gaze of the participant who used the recording equipment during the discussion. In addition, we connected a speakerphone (Anker PowerConf A3301[5]). We assigned speaker IDs A, B, C, and D to the participants in each discussion. Speaker A used the recording PC. Speakers B, C, and D used their own devices.

Fig. 2 shows an example of the recording environment. Speaker A appears in the upper-right corner of the screen. Likewise, speakers B, C, and D appear on the screen.

3.2. Discussion Task

We designed the discussion task and the recording procedure based on the Kyutech Corpus. Each discussion included four participants. The participants played the role of managers of a shopping mall in a fictional city. Their task was to select one restaurant from three candidates after one tenant restaurant closed. Each participant received ten pages of materials before the discussion. The materials included various types of information, such as:

- the candidate restaurants
- the restaurant that had closed and the reason for the closure
- the existing restaurants
- the location of the shopping mall

[3] https://www.elecom.co.jp/products/UCAM-CX80FBBK.html

[4] https://gaming.tobii.com/

[5] https://www.ankerjapan.com/products/a3301

Table 1: tags for utterances.

tag	tag description
(F)	filler
(D)	repair
(Q)	question
(?)	uncertain transcription
(L)	whispered speech
<Laugh>	laugh

- the distribution of visitors by time and gender
- the population of the city
- information about neighboring municipalities

The participants could refer to these materials during the discussion.
We prepared 10 discussion scenarios. The information about the candidate restaurants differed across the scenarios.
The discussion procedure was as follows. First, the participants read the materials for 10 minutes. Next, they discussed the task for approximately 20 minutes. After 20 minutes, we informed the participants of the end of the discussion by ringing a bell. However, the participants could finish the discussion before the bell. They could also continue the discussion after the bell.
We recruited 18 university students[6] who were familiar with each other. We selected four participants for each discussion. A participant could join multiple discussions. However, no participant discussed the same scenario more than once.
After each discussion, the participants completed a questionnaire. The questionnaire included items such as "Which participant controlled the discussion?"

3.3. Annotation

We manually transcribed the recorded discussions using ELAN[7]. We followed the transcription rules of "The Corpus of Spontaneous Japanese(CSJ)" (Maekawa, 2006).
We divided the audio into transcription units based on silent intervals of 0.2 seconds or longer. Each transcription unit corresponded to one line of the transcript. We also assigned tags to each utterance when necessary. Table 1 lists the tags.
The CSJ transcription rules sometimes divide one utterance into multiple transcription units. Therefore, we annotated utterance boundaries in the same manner as the Kyutech Corpus. First, we assigned the symbol "+" to the end of every transcription unit. Next, we replaced "+" with "/" when the interval before the next transcription unit is one second or longer and one of the following conditions is satisfied.

- The transcription unit ends with "a verb" or "an auxiliary verb" + "a conjunction".

[6]They are undergraduate students and graduate students at Kyushu Institute of Technology.

[7]https://archive.mpi.nl/tla/elan

ID	Start	End	Utterance	Gaze_A
A	00:08.410	00:08.540	Yeah+	X
A	00:08.780	00:11.443	Well then, let's get started. Thank you all in advance/	A
B	00:12.433	00:13.052	Thank you/	C
C	00:12.794	00:13.423	Thank you/	C
D	00:13.371	00:14.649	Thank you/	D
A	00:17.794	00:19.220	How should we proceed(Q)/	C
C	00:20.000	00:21.860	(F Well,) why don't we start with A/	C

Figure 3: An example of the corpus data.

Table 2: Corpus Statistics

Measure	face-to-face	remote
Utterance density	0.38	0.40
Non-speech interval rate	0.45	0.44
Speaker transition rate	0.54	0.30

- The next utterance by the same speaker begins with "a conjunction", "a filler", or "an adverb".

In some cases, these rules do not identify the end of an utterance correctly. When we determined that a transcription unit is the end of an utterance, we replace "+" with "*".
We manually annotated whom speaker A was looking at based on the gaze data recorded by the eye tracker. The method of gaze annotation is described in the Appendix.
As a result, we obtained transcription data from 20 discussions. The data include speaker IDs, utterance start and end times, transcription tags, and gaze information. The corpus contains 11,154 utterances. Fig. 3 shows an example of the dialogue data.

3.4. Corpus Statistics

To clarify the characteristics of the Kyutech Remote Corpus, we compared its statistics with those of the Kyutech Corpus (Yamamura et al., 2016). We compared the following three measures.

- Utterance density: the ratio of the total number of utterances to the total discussion time.
- Non-speech interval rate: the ratio of the total duration of non-speech intervals to the total discussion time.
- Speaker transition rate: the ratio of speaker transitions to the total number of utterances.

Utterance density represents the frequency of utterances during a discussion. The non-speech interval rate represents the proportion of time during which no participant is speaking. The speaker transition rate also represents how frequently the speaker changes during a discussion.
Table 2 summarizes the average values for the two corpora. The two corpora showed similar values for utterance density and the ratio of non-speech intervals. In contrast, the Kyutech Corpus showed a speaker transition rate of 0.54, whereas the Kyutech Remote Corpus showed a value of 0.30. Therefore, the speaker transition rate differed between the two corpora. This result suggests that participants tend to continue speaking for longer periods in remote discussions than in face-to-face discussions.

#Task Description
You are an expert in conversation analysis and discussion structure.
You will be given a discussion conducted for the purpose of deciding which restaurant should open a new store in the restaurant area of a shopping mall.
Based on the discussion, analyze which speaker functioned as the facilitator. Then, identify the speaker who functioned most as the facilitator and provide specific reasons for your judgment.
If you determine that multiple speakers functioned as facilitators, you may identify more than one speaker.

#Output Format
Please output your answer in the following format.
Estimation:
A: <Analysis of Speaker A>
B: <Analysis of Speaker B>
C: <Analysis of Speaker C>
D: <Analysis of Speaker D>
Facilitator: <Speaker ID(s) (1–4 speakers)>
Reason: <Reason for the estimation>

#Discussion Data
The following are the utterances from a discussion involving four participants.
(Start Time) - (End Time) [(Speaker ID)]: (Utterance)
⋮

Figure 4: Prompt for LLM.

4. Method

One of the goals of this study is to analyze facilitator characteristics in face-to-face and remote discussions. This section describes the datasets, the definition of facilitators, the facilitator identification method, and the procedure for analyzing facilitator characteristics.

4.1. Dataset

We used the Kyutech Remote Corpus and the Kyutech Corpus in this study. The main difference is as follows:

- The Kyutech Remote Corpus: remote discussions.
- The Kyutech Corpus: face-to-face discussions.

Both corpora target the same decision-making task with four participants, namely multi-party. Therefore, we could compare facilitator behavior between face-to-face and remote discussions under similar experimental conditions.

4.2. Facilitator Definition

This study defined a facilitator as the participant who was considered to have managed the discussion most effectively. We used the participants' answers to the questionnaire item "Which participant played the role of the facilitator? The participant who received the largest number of votes was regarded as the facilitator. If multiple participants received the same largest number of votes, all of them were regarded as facilitators. As a result, in the Kyutech Remote Corpus, 22 of 80 participants were facilitators. Similarly, in the Kyutech Corpus, 10 of 36 participants were facilitators. We used these labels as the ground-truth labels throughout this study.

4.3. LLM-based Facilitator Identification

We used GPT-4o to identify facilitators from discussion transcripts. Each transcript was provided to the LLM together with a prompt. Fig. 4 shows the content of the prompt. The prompt instructed the LLM to identify the facilitator and explain the reason for the decision. The output consisted of the identified facilitator and a textual explanation of the decision.
We set the temperature to 0.0. The LLM analyzed each discussion only once. The LLM was allowed to identify multiple facilitators for each discussion. The explanations generated by the LLM were used in the subsequent analysis only when the identified facilitators matched the ground-truth facilitators.

Table 3: The performance of facilitator estimation

	face-to-face	remote
Accuracy	0.90 (9/10)	0.82 (18/22)

4.4. Feature Analysis Procedure

We analyzed facilitator characteristics based on the explanations generated by the LLM. First, the first author examined every explanation. Next, the first author assigned one or more categories to each explanation. After the categorization, we counted the frequency of each category separately for the face-to-face and remote discussions. We then calculated the proportion of each category with respect to the total number of explanations in each discussion format. Finally, we compared the distributions of the categories between the two discussion formats. This comparison clarifies similarities and differences in facilitator behavior.

5. Results

5.1. Identification Performance

We evaluated the facilitator identification performance of the LLM. We compared the identified facilitators with the ground-truth facilitators. Table 3 summarizes the identification accuracy for the two corpora. Here, we calculated the identification accuracy as the ratio of correctly identified facilitators to the total number of ground-truth facilitators. The LLM correctly identified 9 of 10 ground-truth facilitators in face-to-face discussions and 18 of 22 ground-truth facilitators in remote discussions. We analyzed only the explanations generated for correctly identified facilitators in the following experiments.

5.2. LLM-based Feature Analysis

We analyzed the explanations generated by the LLM to investigate facilitator characteristics. First, the first author classified the explanations into predefined categories. The definitions of each category were as follows.

- Management: Descriptions indicating that the participant manages the discussion or proposes how to proceed.
- Topic Clarification: Descriptions indicating that the participant clarifies the topic or organizes the discussion.
- Enhancement: Descriptions indicating that the participant deepens or activates the discussion.
- Opinion Elicitation: Descriptions indicating that the participant elicits opinions from other participants.
- Questioning: Descriptions indicating that the participant asks questions to other participants.

Table 4: The ratio of criteria by dialogue formats

category	face-to-face	remote
Management	55.6 (5/9)	83.3 (15/18)
Topic Clarification	100.0 (9/9)	88.9 (16/18)
Enhancement	44.4 (4/9)	27.8 (5/18)
Opinion Elicitation	88.9 (8/9)	61.1 (11/18)
Questioning	22.2 (2/9)	16.7 (3/18)
Active Participation	22.2 (2/9)	0.0 (0/18)
Coordination	22.2 (2/9)	22.2 (4/18)
Consensus Building	33.3 (3/9)	61.1 (11/18)

- Active Participation: Descriptions indicating that the participant actively contributes to the discussion or speaks frequently.
- Coordination: Descriptions indicating that the participant accepts other participants' opinions or mediates the discussion.
- Contribution to Consensus Building: Descriptions indicating that the participant summarizes opinions or leads the discussion toward consensus.

Next, we counted the occurrences of each category for the two discussion formats. We then calculated the proportion of each category with respect to the total number of explanations. Table 4 summarizes the frequency of each category in the face-to-face and remote discussions.

According to Table 4, "Topic Clarification" appeared most frequently in both discussion formats. In the remote discussions, "Management" and "Contribution of Consensus" appeared more frequently. However, "Enhancement" and "Opinion Elicitation" appeared less frequently than in face-to-face discussions.

5.3. Validation of LLM-generated Explanations

We evaluated the validity of the explanations generated by the LLM. For each discussion, two evaluators[8] independently reviewed all explanations. Specifically, we provided each evaluator with the explanation generated by the LLM and the corresponding discussion transcript. The evaluators determined which participant (A, B, C, or D) each explanation referred to.

We calculated the agreement rate between the two evaluators for all explanations. The agreement rate was 0.634. On the other hand, we also evaluated the explanations generated for the correctly identified facilitators. The agreement rate was 0.906. These results indicate that the explanations generated by the LLM were generally consistent with the participants' behaviors in the discussion transcripts.

6. Discussion

6.1. Differences in Facilitator Behavior between the Two Discussion Formats

The results showed that facilitator behavior differed between face-to-face and remote discussions. According to the explanations generated by the LLM, "Management" and "Contribution to Consensus Building" appeared more frequently in remote discussions than in face-to-face discussions. Participants in remote discussions often experience delays in communication. They also receive fewer nonverbal cues from other participants in remote discussions. Facilitators may therefore need to manage the discussion more explicitly.

In contrast, "Enhancement" and "Opinion Elicitation" appeared less frequently in remote discussions than in face-to-face discussions. Participants may express their opinions less spontaneously in remote discussions than in face-to-face discussions. As a result, facilitators may spend less time encouraging additional opinions. Therefore, in-depth topic exploration may occur less frequently in remote discussions. These results suggest that remote facilitators emphasize discussion management and consensus building more strongly than face-to-face facilitators.

6.2. Limitations

This study has several limitations. First, there is a difference in the discussion numbers of the Kyutech Remote Corpus and the Kyutech Corpus. Future studies should include a more balanced number of discussions for more reliable comparisons.

Second, only the first author categorized the explanations. However, the reliability of the categorization should be improved by involving multiple annotators.

Finally, this study focuses on discussions conducted in fully remote environments. However, many organizations adopt hybrid meetings in practice, where some participants are co-located and others participate remotely. Facilitator behavior may differ in such environments because facilitators need to coordinate interactions between co-located and remote participants. Therefore, it is useful to construct a hybrid discussion corpus and investigate facilitator behavior in hybrid discussions.

7. Conclusion

This study investigated facilitator behavior in Japanese multi-party remote discussions. We constructed the Kyutech Remote Corpus, a freely available Japanese corpus of multi-party remote discussions.

We analyzed facilitator characteristics using an LLM and compared facilitator behavior between face-to-face and remote discussions under the same discussion task. The Kyutech Remote Corpus contains 20 multi-party decision-making discussions and 11,154 manually transcribed utterances. We used GPT-4o to identify facilitators from discussion transcripts and generate explanations for its decisions. We analyzed the explanations generated by the LLM to investigate facilitator characteristics. The results showed that the LLM identified many ground-truth facilitators correctly. The analysis indicated differences in facilitator behavior between face-to-face and remote discussions. In particular, remote facilitators showed stronger tendencies toward "Management" and "Contribution to Consensus Building" than face-to-face facilitators.

These findings provide useful knowledge for developing facilitation support systems for remote discussions. In future

[8] A total of six evaluators participated in the evaluation. All evaluators had participated in the discussion recordings. However, they did not evaluate discussions in which they had participated.

work, we will balance the number of discussions between two corpora, involve multiple annotators in categorizing the LLM outputs, and construct a hybrid discussion corpus.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 23K11368.

Baiat, G. E., Kirkland, A., and Edlund, J. (2026). Filled and silent pause production in remote, collocated and hybrid meetings. In *Proc. SpeechProsody 2026*, pages 830–833.

Huang, H.-H. and Nishida, T. (2020). Investigation on the fusion of multi-modal and multi-person features in rnns for detecting the functional roles of group discussion participants. In *International Conference on Human-Computer Interaction*, pages 489–503. Springer.

Inaba, M., Chiba, Y., Higashinaka, R., Komatani, K., Miyao, Y., and Nagai, T. (2022). Collection and analysis of travel agency task dialogues with age-diverse speakers. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5759–5767.

Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., et al. (2003). The icsi meeting corpus. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, volume 1, pages I–I. IEEE.

Kirikihira, R. and Shimada, K. (2018). Discussion map with an assistant function for decision-making: A tool for supporting consensus-building. In *International Conference on Collaboration Technologies*, pages 3–18. Springer.

Kraaij, W., Hain, T., Lincoln, M., and Post, W. (2005). The ami meeting corpus. In *Proc. International Conference on Methods and Techniques in Behavioral Research*, pages 1–4.

Maekawa, K. (2006). Construction manual of corpus of spontaneous japanese (in Japanese). Technical report, National Institute for Japanese Language and Linguistics.

Nishimura, R., Iharada, R., Sugamoto, Y., Ishii, Y., Mochizuki, T., and Egi, H. (2023). Discussion support agent system to promote equalization of speech among participants. In *International Conference on Computers in Education*.

Okahisa, T., Tanaka, R., Kodama, T., Huang, Y. J., and Kurohashi, S. (2022). Constructing a culinary interview dialogue corpus with video conferencing tool. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3131–3139.

Onishi, T., Ogushi, A., Tahara, Y., Ishii, R., Fukayama, A., Nakamura, T., and Miyata, A. (2022). A comparison of praising skills in face-to-face and remote dialogues. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5805–5812.

Shiota, T., Yamamura, T., and Shimada, K. (2018). Analysis of facilitators' behaviors in multi-party conversations for constructing a digital facilitator system. In *International Conference on Collaboration Technologies*, pages 145–158. Springer.

Shiota, T., Honda, K., Shimada, K., and Saitoh, T. (2020). Development and application of leader identification model using multimodal information in multi-party conversations. *International Journal of Asian Language Processing*, 30(04):2050019.

Wakita, K. and Shimada, K. (2024). An utterance is enough to the gaze? Gaze detection from utterance information in multiparty discussion. In *2024 International Conference on Activity and Behavior Computing (ABC)*, pages 1–8. IEEE.

Yamamura, T., Shimada, K., and Kawahara, S. (2016). The Kyutech corpus and topic segmentation using a combined method. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pages 95–104.

Appendix

A. Gaze Information Annotation

We manually annotated the gaze target for each utterance in the Kyutech Remote Corpus based on eye-tracker data. We assigned gaze targets only to participant A because only this participant was recorded using the eye tracker. For each utterance, we assigned the ID of the participant that participant A looked at during the utterance. We followed the definition of gaze targets used in the previous study (Wakita and Shimada, 2024). In this study, "a participant looked at another participant" means that the participant looked at the other participant at least once during an utterance. If participant A looked at multiple participants during an utterance, we assigned the ID of the participant with the longest gaze duration.

Participant gaze labels were assigned to each utterance based on the following conditions.

- Label A: participant A looks at their own video image on the screen
- Label B: participant A looks at participant B
- Label C: participant A looks at participant C
- Label D: participant A looks at participant D
- Label X: participant A does not look at any participant