Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4088–4095
Marseille, 20-25 June 2022

Dialogue Corpus Construction Considering Modality and Social Relationships in Building Common Ground

Yuki Furuya[†], Koki Saito[†], Kosuke Ogura[†], Koh Mitsuda[‡], Ryuichiro Higashinaka[‡], and Kazunori Takashio[†]
[†]Keio University, Japan
[‡]NTT Corporation, Japan
[†]{furuyan, t19340ks, t19177ko, kazu}@sfc.keio.ac.jp
[‡]{koh.mitsuda.td, ryuichiro.higashinaka.tp}@hco.ntt.co.jp

Abstract

Building common ground with users is essential for dialogue agent systems and robots to interact naturally with people. While a few previous studies have investigated the process of building common ground in human-human dialogue, most of them have been conducted on the basis of text chat. In this study, we constructed a dialogue corpus to investigate the process of building common ground with a particular focus on the modality of dialogue and the social relationship between the participants in the process of building common ground, which are important but have not been investigated in the previous work. The results of our analysis suggest that adding the modality or developing the relationship between workers speeds up the building of common ground. Specifically, regarding the modality, the presence of video rather than only audio may unconsciously facilitate work, and as for the relationship, it is easier to convey information about emotions and turn-taking among friends than in first meetings. These findings and the corpus should prove useful for developing a system to support remote communication.

Keywords: dialogue, corpus construction, voice/video chat, social relationship, collaborative task, common ground

1. Introduction

In dialogue, information such as knowledge and beliefs that is shared among the participants is called common ground (Clark, 1996; Traum, 1994). Building common ground with the user in dialogue is essential for implementing dialogue agent systems and robots that perform natural interaction with humans.

Research that explicitly considers the common ground has been reported (Lee et al., 2011; Kiesler, 2005; Moratz and Tenbrink, 2006; Chai et al., 2017). These studies have analyzed the common ground in dialogues where workers collaboratively accomplish tasks. In addition, some studies have examined the process of building common ground in text chat (Udagawa and Aizawa, 2019; Mitsuda et al., 2022; Bara et al., 2021).

Although nonverbal cues and relationships between interlocutors are essential in the process of common ground building, they have not been analyzed in depth. As nonverbal behavior between interlocutors is essential for cooperative dialogue, nonverbal cues need to be taken into account (Carney and Harrigan, 2003). For example, nonverbal behaviors such as shaking one's head without speaking can have a facilitating effect on the dialogue with the other party (Kleinke, 1986). The difference in modality between voice and video calls leads to a difference in the results of collaborative tasks through dialogue (Tomprou et al., 2021). By extending text chat to voice chat and video chat, interlocutors can smoothly build common ground based on the information in the other person's voice and facial expressions.

It is also crucial to consider social relationships, as the nonverbal behavior within a dialogue changes depending on the relationship between the interlocutors (Taylor, 1968). Various studies on communication and psychology have investigated the relationship between modality and the interlocutor in the past. In particular, research focusing on communication in video chat has shown that the content and outcome of dialogue varies depending on the social relationships of the speakers (Postmes et al., 2001). A multimodal corpus featuring groups with varying degrees of relationship between the interlocutors has also been collected (Yamazaki et al., 2020).

In this study, we construct a corpus to analyze the influence of the relationship between the interlocutor and the difference in the modality on the process of building common ground. We will extend the previous work on the quantitative analysis of common ground conducted in text chat in order to investigate the effects of differences in communicated modality and social relationships on the common ground building process. We collected dialogues in which a pair of workers performed a cooperative task (proposed in the previous study) in four different conditions: Voice or Voice+Video as a modality, and First meet or Friend as a social relationship. We further investigated how each condition affects the process of building common ground.

Our findings showed that adding the modality or developing the relationship between workers speeds up the building of common ground. Specifically, regarding the modality, the presence of video rather than only audio may unconsciously facilitate work. As for the relationship, it is easier to convey information about emotions and turn-taking among friends than in first meetings. These findings, as well as the corpus, should prove useful for developing a system to support remote

communication.
Section 2 of this paper provides an overview of the data collection process and the characteristics of the collected data, while Section 3 presents a comparative study of the collected data in each condition based on the success rate of the task and the results of a questionnaire. In Section 4, we introduce an index (evaluation metric) to quantitatively analyze the process of building common ground and discuss the differences in the typical process of building common ground under each condition.

2. Data Collection

This section gives an overview of the corpus collected in this study. First, we describe the definitions of nonverbal behavior and social relationships as control factors for the task. Next, we provide an overview of the tasks proposed in previous studies, the data collection experiments conducted in this study, and the characteristics of the collected data.

2.1. Controlling Modality and Social Relationship

We describe the control of the modality conditions and the social relationship conditions for collecting the dialogue data in each condition. Under the modality condition, we use the video on/off function of the video chat tool. This means that when the video chat video is turned off (**Voice**), only the voice is heard by the other party. When the video is turned on (**Voice+Video**), the video will reach the other party along with the voice, and facial expressions and nonverbal behavior will be included in the communication. This condition will be used to analyze the change in the process of building common ground with and without visual information.
In the social relationship condition, we introduced a questionnaire to classify pairs that knew each other before the experiment (**Friend**) and pairs that had never met before (**First meet**). This will be used to analyze whether intimacy helps workers to successfully build common ground.

2.2. CommonLayout

In this study, we use a task called "CommonLayout", originally created by Mitsuda et al., in which two individuals work in pairs to decide on the placement of objects through dialogue (Mitsuda et al., 2022). This task was created to analyze the process of building common ground. Specifically, two workers design a layout of objects into a common one through text chat. Participants can see only their own interface and not their partner's. Then, the similarity of the layouts created by the two workers is used to quantify the intermediate common ground through dialogue. This has the advantage of automatically recording the process of building common ground during a dialogue, thus eliminating the costly manual annotation observed in related work (Udagawa and Aizawa, 2019; Bara et al., 2021).

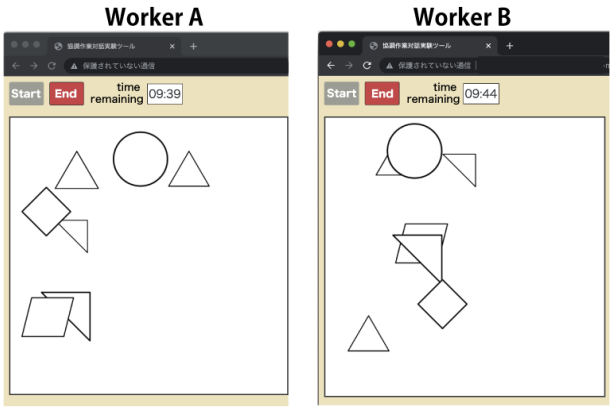

Figure 1: Interface shown to each worker to perform CommonLayout task. Participants can see only their own interface and not their partner's.

The task is performed using a tool in a web browser (Fig. 1). In the task, the objects are initially placed at different locations, and each participant tries to match the placement of the objects by discussing with each other on the video chat tool. At this time, seven objects are displayed on the tool at random locations for each interactor. The participants can move the objects on the screen by dragging the mouse. Once they agree on the placement, they can finish the task by pressing the "End" button at the top of the tool. Figure operations and status changes (such as "Start" and "End") are recorded in a log file with a timestamp.

2.3. Experimental Conditions

The participants were 40 individuals recruited for a fee through a language data collection agency (20 men, 20 women, mean age = 31.8, SD = 11.60). Criteria were that each participant must be an acquaintance (e.g., a friend or family member) who had known each other for at least six months, and first-time pairs were created by swapping the participants. They were permitted to quit the experiment at any time. This study was approved by the ethics committee of Keio University Shonan Fujisawa Campus.
Figure 2 shows the booth space where the participants performed the task. Participants could not hear or see each other directly and were provided with a desk and a comfortable chair. There were two displays on the desk: one showing the task's web browser and one showing the screen of the video chat tool. We used Zoom as the video chat tool for the experimenter's explanation and the conversation with the interlocutor for conducting the task. While conversing, the participants used the mouse to manipulate the task screen and work on the task.
Each participant completed the task and questionnaire in the following flow. After visiting the experimental facility, each participant was briefed on the experiment and data collection and filled out a consent

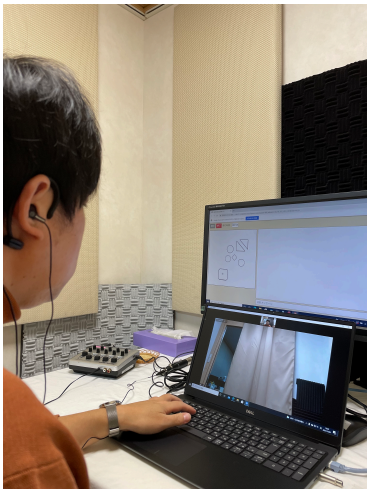

Figure 2: Environment for worker to perform CommonLayout task with a partner who is located in a separate area.

Number of dialogues	80
Amount of time (sec)	30,753
Amount of speech time (sec)	20,659
Number of turns taken	10,387
Number of recorded operations	10,445

Table 1: Statistics of collected data.

form. At the end of the second dialogue, the pairs were switched. Four tasks were conducted with different modalities and social relationship conditions. The order of the conditions was randomized and counterbalanced. Participants were asked to complete a pre-questionnaire, the Japanese version of the Short Big 5 Scale (TIP-J) (Oshio et al., 2012), to provide information about their attributes and personality traits. After each task, another questionnaire was administered to evaluate the participants' impressions of the conversation. The questionnaire consisted of four questions (Q1 to Q4; see below), and the responses were collected on a seven-point Likert scale. The questionnaire was designed to evaluate the participants' impressions of their understanding of the other person's speech, readability of emotions, smoothness of the conversation, and confidence in the success of the task.

Q1 I could understand what my partner was trying to say.
Q2 I could understand my partner's emotion.
Q3 The conversation was smooth.
Q4 I am confident in the outcome of the task.

After the session, the participants were asked to answer a questionnaire about their relationship with the interlocutor. This questionnaire consisted of 16 questions answered on the seven-point Likert scale, referring to previous research on interpersonal relationships (Tanno, 2007).

ID	S	Utterance
…	…	…
U_{82}	A	あと台形ですよね。 (And the trapezoid, right?)
U_{83}	B	はい。台形が、うーんと。 (Yes. The trapezoid is um...)
U_{84}	A	台形、三角形の上とかどうですか？ (The trapezoid... How about putting it on top of the triangle?)
U_{85}	B	あ、このさん、さらにこの小さい三角形の上？ (Oh, this tri..., further up this little triangle?)
U_{86}	A	あ、そうですね。 (Oh, yes.)
U_{87}	B	あ、乗せちゃう <o> のせ、乗せちゃう感じですね。</o> 感じですよねー？ こう、こうあって、こう、こう？ (Oh, we will put the triangle, <o>Pu..., put it, right.</o>right? Like, it is like, like, like...)
U_{88}	A	あ、そうです、そうです、三角三角台形。 (Oh, yes yes, triangle, triangle, trapezoid.)
U_{89}	B	はいはいはいはい。三角、小さい三角、台形。 (Yes, yes, yes, yes. Triangle, little triangle, trapezoid.)
U_{90}	A	そうですね。 (That's right.)
…	…	…

Table 2: Example of multimodal CommonLayout corpus (ID: Utterance ID, S: Subjects, <o>: Utterance overlaps).

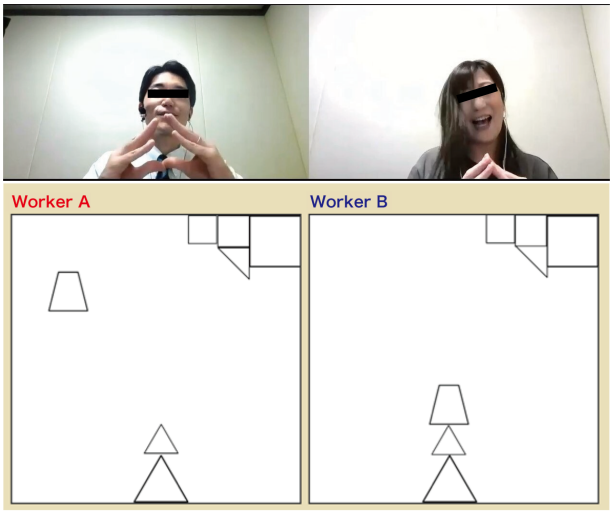

Figure 3: Participants using gestures in multimodal CommonLayout corpus.

2.4. Collected Data

Table 1 shows an overview of the collected data, which consists of 80 dialogues totaling about 513 minutes. There are 20 dialogues for each condition of modality and social relationship. The number of recorded operations is the number of times an object is manipulated in a task. The number of object manipulations is approximately the same as the number of turns taken.
An example of the dialogue that took place at the end of the task is shown in Table 2. In U_{82}, both sides con-

	Voice		Voice+Video		
	First meet	Friend	First meet	Friend	p-value
Number of conversations (available log data)	20 (20)	20 (20)	20 (20)	20 (19)	–
Average number of operations	134.15	130.85	136.50	122.63	0.94
Number of turns taken	124.80	134.30	120.95	143.58	0.48
Average total speech time (sec)	256.18	264.16	246.30	268.67	0.94
Average task time (sec)	388.28	353.46	393.26	403.43	0.80

Table 3: Statistics of operations and utterances.

firm that the trapezoid is the final object whose position has not been determined. In U_{87} and U_{88}, they are using gestures to confirm the order of the objects, including the placement of the objects determined so far (Fig.3). In U_{90}, they make a statement that is agreed upon by the other party, and we can see that the final common ground has been established. In addition to this dialogue, other dialogues using gestures etc. were confirmed under the Voice+Video condition.

The collected data can be summarized as follows. Note that one task log data was not included in the analysis because it had a problem with the experimental tool[1].

Recorded audio 2-ch audio with each speaker's speech split into left and right (80 items)

Recorded video Voice+Video condition only (40 items)

Log data Log data of object movement (79 items)

Questionnaire after each task Impression ratings during the task (80 items)

Post-experiment questionnaire Relationship between the pairs of interlocutors (40 pairs)

3. Analysis of Collected Corpus

In this section, we describe the results of our analysis regarding the effects of each modality condition and each social relationship condition on the common ground building process by comparing the statistics of the collected data.

3.1. Statistics of Operations and Utterances

The statistics of the collected data classified by each condition are shown in Table 3. The number of operations indicates the total number of times the figures were manipulated in the task among the participants. The time to complete the task is the average number of seconds between when the "Start" button was pressed and the "End" button was pressed in the task tool. The number of turns taken is the count of turns taken, and the average total speech time is the average of the total speech time of each speaker. The number of turns taken and the total speech time were analyzed using a speech analysis program written in Python(Bechtold and Geier, 2015).

[1]One user's operation was not recorded at all.

The table indicates that, regardless of the modality, Friends had more turn-taking and fewer object manipulations than First meet. This is because the pair of friends tended to exchange information more smoothly and move the objects more efficiently. Note that ANOVA was performed on each of these indices to analyze any differences between the four conditions. P-values are listed in the p-value column, and no significant differences were found for any condition or index.

3.2. Success and Failure of the Task

To compare the final rate of accomplishment of the task, i.e., the rate of agreement of the figures, we classified the final object placement. The following classification was performed in the same way as in previous studies on text chat (Mitsuda et al., 2022).

1. **Perfect** Placements of all objects are an exact match
2. **Shifted** The original position of the whole figure is different
3. **Resized** The whole graphic has a different size
4. **Symmetric** The position of some objects is different by contrast
5. **Switched 1** Different positions of the same object
6. **Switched 2** Different objects in different positions
7. **Scattered** Objects in completely different positions

As in the previous study, the participants were not instructed to align the position of the entire figure (origin). Therefore, in this study, common ground among the participants is considered to be completely established in the patterns of "Perfect" and "Shifted".

Table 4 shows the corresponding number of each pattern for the collected dialogues. The relationship between each object placement pattern and each condition was analyzed by conducting an ANOVA. There was no significant difference between the conditions in terms of the placement pattern. Participants succeeded with the task for a high percentage regardless of the modality condition (e.g., Voice or Voice+Video) or the social relationship condition (e.g., First meet or Friend). When we sampled the dialogues classified as "Scattered", we found that there were cases in which the participants noticed a gap in their perceptions at the end of the task but were unable to correct it because they ran out of time, and cases in which they proceeded with the task without matching the detailed images.

		Voice		Voice+Video		
		First meet	Friend	First meet	Friend	p-value
Success rate of task		60.00%	55.00%	60.00%	57.89%	0.40
Success	Perfect	60% (12)	55% (11)	50% (10)	58% (11)	–
	Shifted	0% (0)	0% (0)	0% (0)	0% (0)	–
Failure	Symmetric	0% (0)	0% (0)	0% (0)	0% (0)	–
	Resized	5% (1)	20% (4)	20%(4)	16% (3)	–
	Switched 1	0% (0)	0% (0)	5% (1)	0% (0)	–
	Switched 2	0% (0)	5% (1)	0% (0)	0% (0)	–
	Complete mismatch	35% (7)	20% (4)	15% (3)	26% (5)	–

Table 4: Statistics of task results.

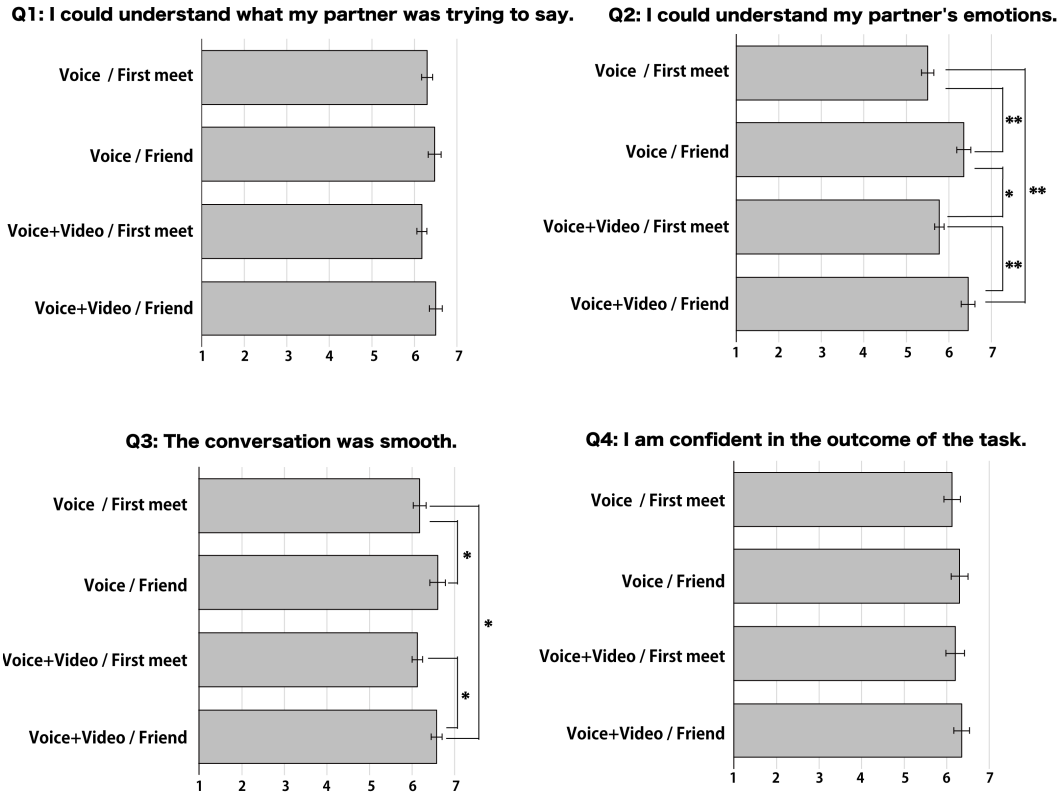

Figure 4: Average questionnaire scores.
(** represents $p < .01$, * represents $p < .05$, error bars show SE)

3.3. Questionnaire After Each Task

Figure 4 shows the results of each questionnaire. To investigate the influence of the impressions received by the participants in each condition on the process of building common ground, we compared the results of the questionnaire administered after each task for each condition. The Steel-Dwass multiple comparison method (Dwass, 1960) was used to investigate whether there were significant differences.

First, the results of the summary questionnaire Q4 show that there is no difference in confidence for each condition. This is consistent with the fact that there is no difference in the success rate of the task for each condition shown in Table 4, indicating that the participants were able to estimate the task results correctly.

Next, we discuss Q1 to Q3 regarding the understanding of utterances and dialogue. Q1 shows that there is no difference in the basic understanding of the content. However, there is a significant difference in the readability of emotions and the smoothness of the dialogue between First meet and Friends. There is no difference in the modality between voice and video. In other words, there is no clear difference between the two modalities in terms of emotion readability and smoothness of dialogue, unlike the two relationships.

The success rate of the task shows that even under the Voice / First meet condition, the participants succeeded in the task to the same extent as under the other conditions. This suggests that even if it was difficult to understand the information about emotions and turns,

the same level of working results could be achieved by going through different processes of building the common ground.

4. Analysis of the Grounding Process

In this section, we further investigate the differences in the process of building common ground caused by different conditions of modalities and social relationships. In the following, we describe the results of quantifying the common ground and the results of applying time-series clustering, following the previous study (Mitsuda et al., 2022).

4.1. Quantification of Common Ground

In order to clarify the differences in the process of building common ground between the conditions, where the task was successful even if the pair was new to each other or interacted only by voice, we analyzed the common ground using quantitative indicators. We focus on the distance between figure placements, which is the sum of the distances of the differences of the vectors defined between any two objects in the task. In a previous study in which the same task was conducted in a text chat (Mitsuda et al., 2022), the data on the relationship between the distance between object placements and the number of each step was classified using clustering methods, and it was found that there were several patterns in the construction of common ground. By plotting the distance between the object placements for each time step, we can measure the state of the common ground at each one. This value is lower for object placement matches between workers, i.e., when the common ground has been established.

Figure 5 shows the average distance between the object placements in successful patterns (patterns that resulted in "Perfect" or "Shifted") for each condition. The y-axis of the graph represents the distance between the object placements and the x-axis is the number of steps in the object operation. The number of steps of the object manipulation was normalized to 100 using a linear transformation. In all conditions, the distance between the object arrangements converges to a smaller value near the final 100 steps, indicating that the object arrangements match between workers. The results show that the initial building of common ground stagnates in the case of Voice/First meet (purple), while in the case of Voice+Video/Friends (blue), it proceeds smoothly from start to finish. Under the First meet condition, there is a lot of confirmatory action at the beginning along with moving objects while exploring the other person's intentions, and the building of common ground in the early part of the task becomes loose. Comparing the conditions, it is easier to build common ground with Friends than with First meet. In the case of voice and video, the work progresses more easily with video.

To summarize the conclusions based on these tendencies, adding the modality or developing the relationship has a positive effect on building common ground. As for the modalities, it is less likely for building common ground to stagnate when videos are available. The results of the questionnaire show that the workers are not aware of this difference, which suggests that they may be unconsciously communicating various information to each other through video. As for the relationship, it becomes easier to read emotions and know the timing of taking turns, and building common ground proceeds smoothly.

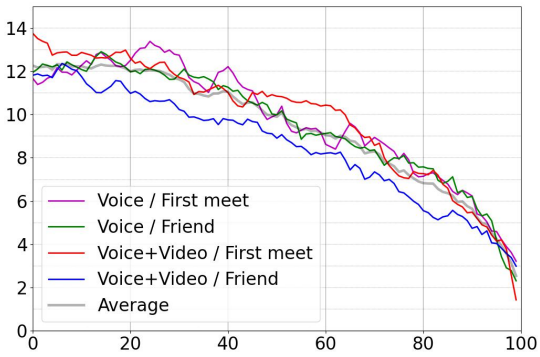

Figure 5: Average distance between object placements of successful patterns in each condition.

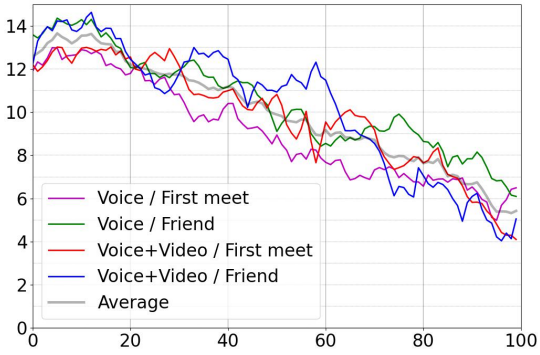

Figure 6: Average distance between object placements of failure patterns in each condition.

Figure 6 shows the average distance between object placements in the failed pattern. The distance increases in all conditions at first and then does not converge as much as the successful pattern. Unlike the case of success, there is a section where Voice+Video/Friends has the largest distance. The turn-taking seems smoother with Voice+Video/Friends, which may make it easier for misunderstandings to occur. We observed cases where the participants were proceeding with the task based on their own assumptions in Voice+Video/Friends, resulting in a gap between their perceptions. In addition, if the goal image is not shared, as in the case of success, it will not

be complete even if the distance is reduced. For example, cases in which participants proceeded with the task without matching the detailed images were observed.

4.2. Clustering

We also used a clustering method to clarify the typical process of building common ground to validate the differences described above. Following the method applied in the previous study (Mitsuda et al., 2022), we clustered the data for 79 dialogues that were successfully logged by using k-Shape (Paparrizos and Gravano, 2015), a common method for clustering time-series data.

Figure 7 shows the results of clustering created by k-Shape. When the number of clusters was further increased, similar clusters were observed; thus, we classified the data into five clusters. The y-axis is the distance between object placements and the x-axis represents the number of steps after normalizing the number of object operations to 100 using a linear transformation. To apply k-Shape, the distances between the objects were normalized to follow a normal distribution with mean = 0 and variance = 1; thus, it was approximately -2 to 2.

We sampled several conversations classified into each cluster and compared their dialogue content. Cluster 1 is the pattern in which the common ground building progresses most steadily and averagely, while Clusters 2–4 are the patterns in which the common ground building progresses significantly after the middle stage. Cluster 5 is the pattern in which the construction of the common ground progresses until about 40 steps in the middle stage but then stagnates, and the construction of the common ground is carried out in the latter half. In this pattern, there were cases where the participants were discussing the graphic pattern of the overall image again because they noticed differences in their persistence and mutual recognition when they started to adjust the detailed graphic positions. The percentage of each cluster for each modality condition and social relationship condition is shown in Table 5. A total of 58% of the clusters under the Voice+Video/Friend condition were classified as Cluster 3, which is smooth common ground building. The results are consistent with the discussion in Section 4.1; i.e., it is less likely for building common ground to stagnate when videos are available, and as for the relationship, building common ground proceeds smoothly.

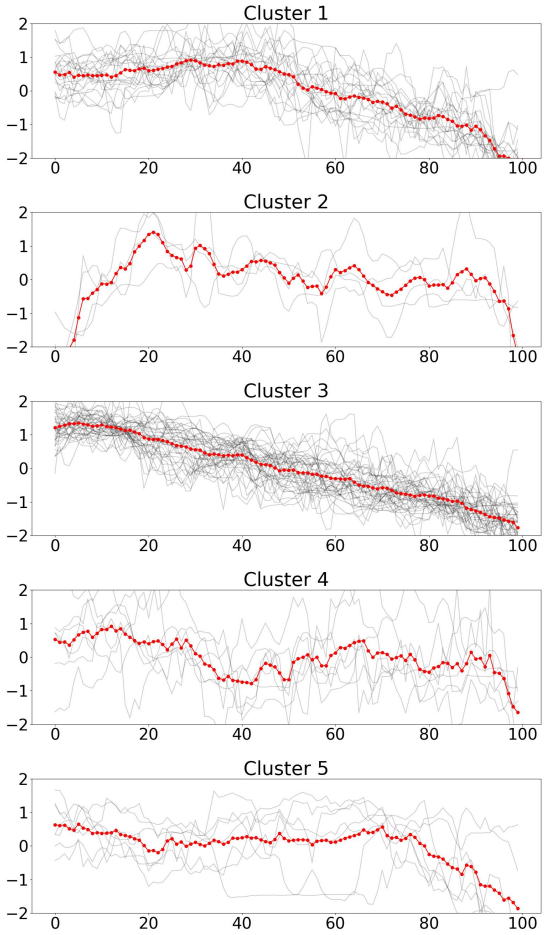

Figure 7: Clustering results.

5. Summary and Future Work

In this study, we controlled multiple modality conditions (Voice+Video/Voice) and social relationship conditions (First meet/Friend) and collected dialogues in which participants engaged in the "CommonLayout" task using a video chat tool.

To quantitatively analyze the process of building common ground, we used the distance between object placement as an indicator to observe the process of building common ground in detail. Comparing the conditions revealed that it is easier to build common ground with Friends than with First meet. In the case of audio and video, the work progresses more easily with video. While the process of building common ground is different, participants adjust well to the final building of the common ground, and succeed with the task for a high percentage.

In this work, we investigated the process of building common ground on the basis of the similarity of layouts and the questionnaire results. However, it will be important to confirm the validity of the results by conducting further analysis based on the contents of dialogue and video. The temporal density of graphic manipulations and their relationship to each utterance will be analyzed through the tagging of utterances and manipulations. In addition, by analyzing the corpus collected in this study in more detail, including the difference from the data collected by text chat, we will analyze how each modality explicitly affects the building of common ground. Under the Voice+Video condition, since the video is recorded, it is possible to annotate gaze and emotion-related actions. We will analyze why the process of building common ground becomes

	Voice				Voice+Video			
	First meet		Friend		First meet		Friend	
Cluster 1	40%	(8)	20%	(4)	20%	(4)	26%	(5)
Cluster 2	5%	(1)	5%	(1)	5%	(1)	5%	(1)
Cluster 3	40%	(8)	45%	(9)	60%	(12)	58%	(11)
Cluster 4	5%	(1)	10%	(2)	10%	(2)	11%	(2)
Cluster 5	10%	(2)	20%	(4)	5%	(1)	0%	(0)

Table 5: Percentage of each cluster accounting for each condition.

smoother by focusing on the propagation of emotions under the Video condition. This study's findings should also prove useful for understanding the dialogue process in video chat and for implementing a system to support remote communication.

6. References

Bara, C.-P., CH-Wang, S., and Chai, J. (2021). MindCraft: Theory of mind modeling for situated dialogue in collaborative tasks. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 1112–1125.

Bechtold, B. and Geier, M. (2015). python-soundfile. `https://github.com/bastibe/python-soundfile`.

Carney, D. R. and Harrigan, J. A. (2003). It takes one to know one: Interpersonal sensitivity is related to accurate assessments of others' interpersonal sensitivity. *Emotion*, 3(2):194–200.

Chai, J. Y., Fang, R., Liu, C., and She, L. (2017). Collaborative language grounding toward situated human-robot dialogue. *AI Magazine*, 37(4):32–45.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Dwass, M. (1960). Some k-sample rank-order tests. *Contributions to probability and statistics*, pages 198–202.

Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, volume 2005, pages 729–734.

Kleinke, C. L. (1986). Gaze and Eye Contact. A Research Review. *Psychological Bulletin*, 100(1):78–100.

Lee, K., Hwang, J. H., Kwon, D. S., and Choo, H. (2011). Bring common ground into robotics. *International Journal of Humanoid Robotics*, 8(3):607–629.

Mitsuda, K., Higashinaka, R., Ohga, Y., and Yoshida, S. (2022). Dialogue Collectionn for Recording the Process of Building Common Ground in a Collaborative Task. In *Proceedings of International Conference on Language Resources and Evaluation*.

Moratz, R. and Tenbrink, T. (2006). Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial Cognition & Computation*, 6(1):63–107.

Oshio, A., Abe, S., and Cutrone, P. (2012). Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J). *The Japanese Journal of Personality*, 21(1):40–52.

Paparrizos, J. and Gravano, L. (2015). K-shape: Efficient and accurate clustering of time series. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, volume 2015-May, pages 1855–1870.

Postmes, T., Spears, R., Sakhel, K., and De Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, 27(10):1243–1254.

Tanno, H. (2007). Interaction Frequency and Functions of Friendship for Undergraduates. *The Japanese Journal of Personality*, 16(1):110–113.

Taylor, D. A. (1968). The Development of Interpersonal Relationships: Social Penetration Processes. *Journal of Social Psychology*, 75(1):79–90.

Tomprou, M., Kim, Y. J., Chikersal, P., Woolley, A. W., and Dabbish, L. A. (2021). Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLOS ONE*, 16(3):e0247655.

Traum, D. R. (1994). A computational theory of grounding in natural language conversation. Technical report, Rochester Univ NY Dept of Computer Science.

Udagawa, T. and Aizawa, A. (2019). A natural language corpus of common grounding under continuous and partially-observable context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7120–7127. AAAI Press.

Yamazaki, Y., Chiba, Y., Nose, T., and Ito, A. (2020). Construction and analysis of a multimodal chat-talk corpus for dialog systems considering interpersonal closeness. In *Proceedings of International Conference on Language Resources and Evaluation*, pages 443–448.