Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 4060–4068
Marseille, 20-25 June 2022

Data Collection for Empirically Determining the Necessary Information for Smooth Handover in Dialogue

Sanae Yamashita, Ryuichiro Higashinaka
Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan
yamashita.sanae.w7@s.mail.nagoya-u.ac.jp, higashinaka@i.nagoya-u.ac.jp

Abstract

Despite recent advances, dialogue systems still struggle to achieve fully autonomous transactions. Therefore, when a system encounters a problem, human operators need to take over the dialogue to complete the transaction. However, it is unclear what information should be presented to the operator when this handover takes place. In this study, we conducted a data collection experiment in which one of two operators talked to a user and switched with the other operator periodically while exchanging notes when the handovers took place. By examining these notes, it is possible to identify the information necessary for handing over the dialogue. We collected 60 dialogues in which two operators switched periodically while performing chat, consultation, and sales tasks in dialogue. We found that adjacency pairs are a useful representation for recording conversation history. In addition, we found that key-value-pair representation is also useful when there are underlying tasks, such as consultation and sales.

Keywords: dialogue handover, discourse processing, adjacency pairs

1. Introduction

In recent years, due to the advances in deep learning-based techniques, we are seeing the emergence of highly sophisticated dialogue systems (Brown et al., 2020; Pichl et al., 2020; Ou and Lin, 2020; Adiwardana et al., 2020; Roller et al., 2021). Despite this progress, dialogue systems still struggle to achieve fully autonomous transactions with users, necessitating human operators to take over the dialogue to complete the transactions when problems occur. However, it may be difficult for humans to take over the dialogue smoothly because the content of the dialogue may be complex and/or difficult to immediately grasp. Some systems show the raw dialogue history, highlighted keywords (e.g., Kawahara et al. (2021)), or summary (e.g., Yamashita and Higashinaka (2021)), but even so, it is not yet clear what kind of information is most effective for a smooth handover. Here, we define *handover* as joining a dialogue in the middle and achieving the original objective of the dialogue.

In this study, in order to identify the necessary information for a smooth handover of dialogue, we conducted a data collection experiment in which one of two operators talked to a user and switched with the other operator periodically, exchanging notes when the handover took place (Fig. 1). By examining these notes, it may be possible to identify the information necessary for handing over the dialogue.

In our experiment, we collected 60 dialogues in which two operators switched periodically while performing chat, consultation, and sales tasks in dialogue. Our analysis showed that adjacency pairs are a useful representation for recording conversation history. In addition, we found that key-value-pair representation is also useful when there are underlying tasks, such as consultation and sales.

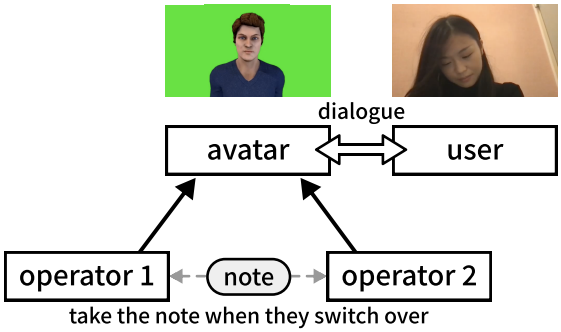


Figure 1: Schematic diagram of dialogue handover experiment. Two operators switch periodically to talk with a user while exchanging notes.

Section 2 of this paper describes related research. In Section 3, we present our dialogue handover experiment consisting of three different dialogue tasks (chat, consultation, and sales) in which two operators take turns talking with a user while taking and passing notes that are useful for the handover, resulting in 60 dialogues/notes with 240 operator switches. In Section 4, we analyze the notes to identify the information necessary to hand over the dialogue.

2. Related Work

The handover in dialogues from systems to operators has been researched in the context of call routing, where a call from a user is transferred to an appropriate operator in a call center. Gorin et al. (1997) developed a system that identifies the topics of a dialogue and routes the call to the appropriate operator. Walker et al. (2000) proposed a method to detect a problematic situation in dialogue by examining the features (including acoustic, natural language understanding, and dia-

logue manager features) of the dialogue so that the call can be handed to an operator at the appropriate timing. However, there has been little research on the actual type of information to be shown to an operator during call routing.

Kawahara et al. (2021) proposed a framework in which a system basically performs most of the dialogue, and if a problematic situation is detected, it nudges a human operator to take over. This makes it possible for one operator to talk to multiple users at the same time. In this system, an operator is shown raw speech recognition results with highlighted keywords (e.g., places and person names) to help him or her grasp the content as easily as possible. Yamashita and Higashinaka (2021) conducted a dialogue handover experiment in which two parties continued a given conversation with one having the whole context and the other only a partial context (i.e., a summary of the dialogue). The aim of that experiment was to determine what kind of summary was effective for continuing the conversation. They reported that presenting the operator with the final lines of a dialogue was the most helpful. Our work differs from these studies in that we empirically examine the necessary information, including summaries, raw text, and keywords, that may be necessary for the handover of dialogue.

For the situation in which multiple speakers take turns to provide dialogue service, Huang et al. (2018) presented Evorus, a human-in-the-loop dialogue system. In Evorus, multiple crowd workers and chatbots decide on the next system utterance from among candidates suggested either automatically or by other crowd workers. Arimoto et al. (2020) conducted a data collection experiment in which multiple operators with different skills participated in a dialogue as one operator and found that operators could learn each other's skills through the dialogue. In our study, as in the above, two operators work together to provide a dialogue service; however, our work differs in that we focus on the process of handover and the information necessary for making it smooth.

We are particularly interested in the presentation of dialogue content to facilitate its understanding by operators. Various representations of dialogue have been proposed, e.g., topic segments (Hearst, 1997; Jiang et al., 2021), discourse segments (Grosz and Sidner, 1986), information states (Larsson and Traum, 2000), and adjacent pairs (Schegloff and Sacks, 1973), along with their extension forms such as comment-response pairs (Stivers, 2013) and sequence organizations consisting of three or more parts (Schegloff, 2007) (e.g., initiation-response-follow-up and initiation-response-approval (Tsui, 1989; Enomoto et al., 2005)). In this study, while keeping these structuring methods in mind, we investigate empirically which representation is most effective for the handover of dialogue.

3. Dialogue Handover Experiment

In order to investigate the information necessary for dialogue handover, we conducted a data collection experiment in which one of two operators talked to a user and switched with the other operator periodically while exchanging notes. By examining such notes, it should be possible to identify the information necessary for handing over the dialogue. In this section, we present the overview and design of the experiment, which includes the dialogue tasks, procedure for operator switching, post-experiment questionnaire with the dialogue participants, and post hoc interview with the operators.

3.1. Overview

A schematic diagram of the dialogue handover experiment is shown in Fig. 1. In this experiment, two operators act as one operator, swapping at regular intervals, and interact with one user. While one operator is talking to the user, the other operator is talking to another user. When the operators switch, they exchange the user they are speaking to. That is, two dialogues are conducted in parallel. This is mainly to make the experiment time-efficient, but this situation, which requires the operators to switch their mindset from the last dialogue to the next dialogue quickly, imposes a significant cognitive load. This hectic situation may encourage them to generate the necessary notes for understanding the dialogue quickly.

We recruited four operators, two women and two men, and put them in pairs consisting of two women (operator pair 1) and two men (operator pair 2). Each operator pair interacted with ten users three times each. There were a total of 20 users, each of whom participated in three dialogue tasks in random order (the tasks are detailed in the next subsection) for a total of 60 dialogues. The experiment was conducted in Japanese.

3.2. Dialogue Tasks

To identify the general elements for a handover of various types of dialogue, we conducted the experiment with multiple dialogue tasks. Specifically, we prepared three dialogue tasks: chat, consultation, and sales. In the chat dialogue, the operator acts mainly as a listener and the user as a speaker. The speaker talks about their favorite things and the listener attentively listens, asking questions and responding as necessary. In the consultation dialogue, the operator acts as a travel agent and the user as a customer. They talk about travel plans for a given destination based on a document listing various tourist spots and accommodations. To make the consultation more realistic, we chose Slovenia as the travel destination, as it is a place most participants had little prior knowledge of. In the sales dialogue, the operator acts as a salesperson and the user as a customer. The salesperson recommends one of three vacuum cleaners, which are shown in Fig. 2.

Item	Statement
Naturalness	The dialogue was natural.
Objective achieved	The objective of the dialogue was smoothly achieved.
Satisfaction	I was satisfied with the dialogue.
Difficulty in understanding the context†	I had difficulty understanding the context at times.
Difficulty with next utterance†	Sometimes I did not know what to say.
Usefulness of other operator's note†	The note written by the other operator was useful to me.
Usefulness of own note†	The note written by myself was useful to the other operator.
Cooperation†	I was able to cooperate smoothly with the other operator.
Content understanding‡	I was able to understand what the operator was saying.
Individuality‡	I was able to get a sense of the operator's personality.
Consistency‡	There was consistency in what the operator was saying.
Confidence‡	I felt that I could trust the operator.
Attention‡	I felt that the operator was paying attention to me.

Table 1: Questionnaire items. A dagger (†) indicates questions posed only to operators and a double dagger (‡) indicates questions posed only to users.

(a) Robot type (b) Stick type (c) Canister type

Figure 2: Vacuum cleaners used in sales task[1].

3.3. Operator Switching

The duration of each dialogue was set to ten minutes. The operators switched at approximately two, four, six, and eight minutes after the start of the dialogue. All dialogues were conducted using Zoom[2], a web conferencing system, with the video on. To prevent the user from detecting the switching of operators, the operators changed their voices by using a voice changing software[3] and appeared as an avatar[4], as depicted in Fig. 1. In addition, the operators were instructed to use a similar register and similar linguistic expressions as much as possible. The users were not informed of the operator switching.

Each operator was instructed to take notes and pass them on to the next operator every time they switched. We did not specify how to take the notes or provide them with any specific instructions, with the idea that they could empirically discover an effective strategy for taking notes through trial and error. Since there was a possibility that graphical information would be required, we asked the operators to take notes on paper (A4 size). Operators talked with the user in soundproof booths close to each other and exchanged the notes in person.

3.4. Questionnaire and Discussion

At the end of each dialogue, the operators and the user filled out questionnaires indicating their level of agreement with 13 statements on a 7-point Likert scale. Table 1 lists the questionnaire items for each role (operator and user).

In addition, after three dialogues (i.e., three dialogue tasks) had been completed with a user, both operators and the user indicated which of the three dialogue tasks were the easiest/hardest to deal with, what had impressed them, and what they had noticed. Operators also answered which dialogue task they found the smoothest/least smooth for cooperating with the other operator.

After each dialogue, the operators discussed ways of improving the note-taking process for ten minutes.

After every six dialogues, the operator pairs were interviewed by the experimenter for about 15 minutes to review the dialogues they had just completed. In the interview, the operators answered which information in the notes was helpful as well as the handover situations in which they felt they had failed.

3.5. Example of Dialogues and Notes

After the experiment, we manually transcribed all utterances, the contents of the notes, the discussions, and the interviews. Figure 3(a)–(c) shows example dialogues for each task: (a) is an example of a chat dialogue, where the operator asked in-depth questions about music, (b) is an example of a consultation dialogue, where the operator asked the user about her intended duration of stay, and (c) is an example of a sales dialogue, where the operator asked the user about his budget. Figures 5 and 6 show examples of notes, where one operator took notes in black ink and the other in red.

[1](a) https://www.irobot-jp.com/product/i3/, (b) https://www.dyson.co.jp/dyson-vacuums/cordless/dyson-micro/dyson-micro-pro.aspx, (c) https://store.miele.co.jp/INTERSHOP/web/WFS/Miele-JP-Site/ja_JP/-/JPY/ViewProduct-Start?SKU=10810670-set1

[2]https://explore.zoom.us/products/meetings/

[3]https://clownfish-translator.com/voicechanger/

[4]https://store.steampowered.com/app/274920/FaceRig/

U	I often listen to songs that were popular in the past.
O	Oh, yeah, that's nice. Do you play that music on your guitar?
U	Oh, I do.
O	That's good.
O	Could you tell me about your favorite artists?
U	Well, I like a band called LUNASEA.
O	Oh, that's good, that's cool.
U	I'm practicing playing LUNASEA songs on the guitar.

(a) Chat

O	How many days are you planning to stay?
U	Ah.
U	I'm not even sure if it's possible to go there for about four days and three nights, but that's about what we're planning on doing.
O	Four days and three nights, right?
U	Yes.
U	What do you think? I was thinking of asking you if I can visit the spots sufficiently in that period. I haven't been able to take a day off from work yet, so...

(b) Consultation

O	Oh, thank you very much. By the way, what kind of budget do you have?
U	Hmmm. Well, let's see. I'd say within 50,000 yen if possible.
O	Within 50,000 yen. Okay.
O	Within 50,000 yen. Yes. OK, let me make some recommendations considering that price.
U	Uh-huh.
O	Let me tell you, the price of Dyson is 49,800 yen. Yes. I think that suits your budget.

(c) Sales

Figure 3: Example of dialogue for each dialogue task, where operator is denoted by O and user by U. No switching of operators occurs in these examples. The above dialogues have been translated from the original Japanese to English by the authors.

	Chat	Consultation	Sales
Speech duration	3 m, 57 s	6 m, 19 s	6 m, 13 s
Utterances	86.1	82.9	86.1
Characters	1312 (15.2)	2175 (25.0)	2057 (23.9)
Words	778.0	1217.3	1210.8
Unique words	177.3	280.2	279.2

(a) Utterances of operators

	Chat	Consultation	Sales
Speech duration	6 m, 26 s	4 m, 21 s	4 m, 39 s
Utterances	71.3	85.5	87.6
Characters	2262 (31.7)	1474 (17.2)	1626 (18.6)
Words	1320.2	883.3	971.3
Unique words	278.2	192.7	212.2

(b) Utterances of users

Table 2: Statistics of utterances. The number in parentheses is the number of characters per utterance.

	Chat	Consultation	Sales
Characters	208.4	115.3	120.5
Words	105.4	57.8	63.6
Unique words	40.7	18.8	21.1
Arrows	151	116	115

Table 3: Statistics of notes.

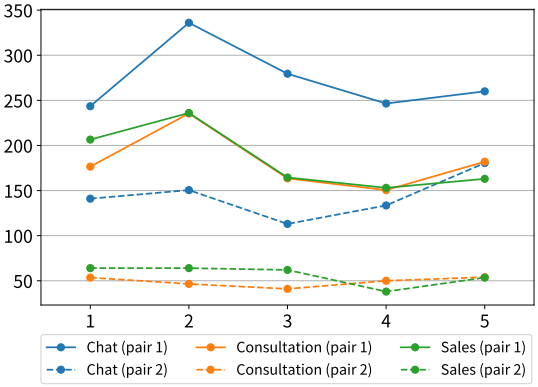


Figure 4: Time transitions of the number of characters in a note. The horizontal axis is the number of times the operators finished two dialogues. An operator pair conducted two dialogues in parallel, switching with each other. The average number of characters in the notes for both dialogues is plotted.

4. Analysis

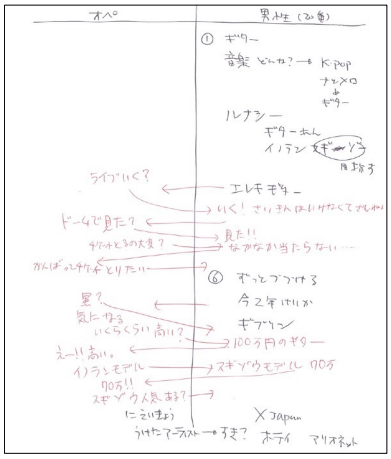
In the following sections, we present our analysis of the dialogue statistics, the notes, the discussions, the interviews, the utterances after switching, and the questionnaire results.

4.1. Dialogue Statistics

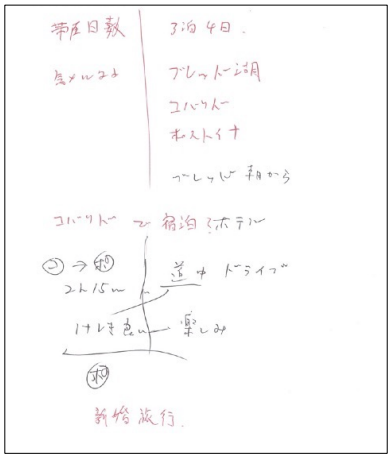
We first examined the statistics of the dialogues. Table 2 lists the dialogue statistics for each speaker role. Over the course of a 10-minute dialogue, the operator made 80 to 90 utterances and the user made 70 to 90. The other statistics were different depending on the task. In the consultation and sales tasks, the number of characters per utterance differed significantly between operators and users, indicating that the operator had taken the initiative in the dialogue. In contrast, in the chat task, the number of characters per utterance of the operator was shorter than that of the user. This indicates that the user was taking control of the dialogue, and the operator mainly provided brief feedback.

4.2. Note Statistics

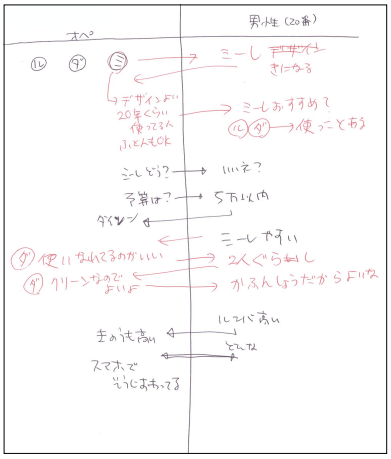
Table 3 shows the number of characters, words, unique words, and arrows in the notes for each dialogue task. The number of characters and words in the notes is roughly proportional to the user's average number of characters per utterance. They increase in the order of consultation, sales, and chat. The number of words in the chat task is higher than in consultation and sales, reflecting the greater variety of topics.



(a) Example of note in chat task. In this dialogue, the user mentions that he likes to play the Gibson guitar he bought. He also likes to listen to music and has been to live concerts. Most of the utterances are connected by arrows, e.g., U: Electric guitar→O: Do you go to their concerts?→U: Yes! Too bad I can't go to concerts these days.→ · · ·

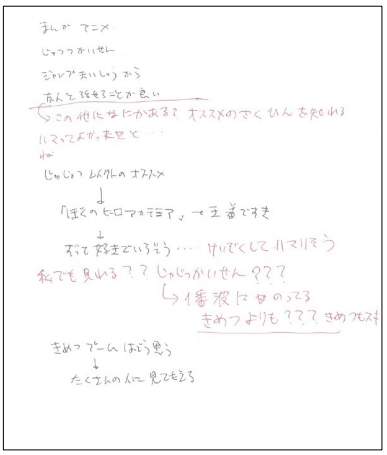


(b) Example of note in consultation task. In this dialogue, the user mentions that she wants to spend four days and three nights. There are key-value pairs, e.g., "Length of stay | Four days and three nights", as well as utterances connected by arrows, e.g., U: I want to drive there.→O: The view is beautiful.→U: I'm looking forward to it.

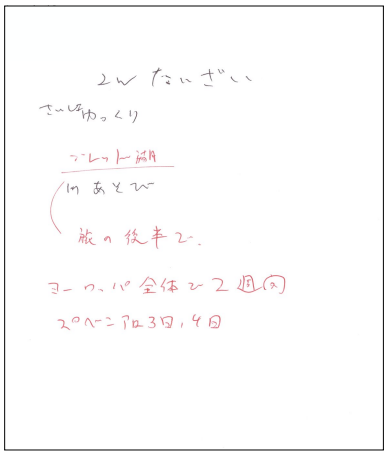


(c) Example of note in sales task. In this dialogue, the user wanted a vacuum cleaner fitting his budget. All utterances are connected by arrows, e.g., U: The robot type is too expensive for me.→O: It is expensive, but high-spec.→U: For example?→ · · ·. A key-value format can be seen in the middle: Budget→Within 50,000 yen.

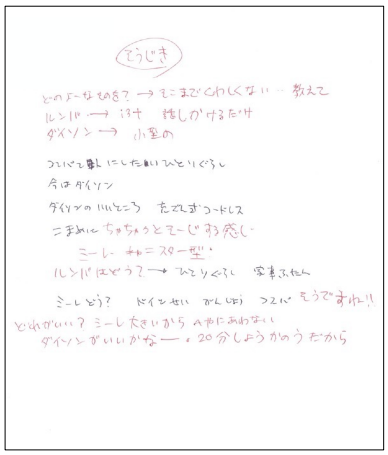
Figure 5: Examples of notes taken at the end of the experiment. In the captions, operator is denoted by O and the user by U. (a) (c) The operator and user utterances are divided into left and right, and arrows connect the utterances. (b) Information is summarized in two formats: key-value pairs and a format similar to that in (a) and (c).



(a) Example of note in chat task.



(b) Example of note in consultation task.



(c) Example of note in sales task.

Figure 6: Examples of notes taken at the beginning of the experiment.

Figure 4 shows the transition of the number of characters in the notes per dialogue. The number of characters written by operator pair 1 peaked in the second dialogues, after which it decreased and tended to converge to approximately 250 characters in chat and 170 characters in consultation and sales. Except for the notes of the first dialogues, the utterances of operators and users were separated into left and right. The number of characters written by operator pair 2 for consultation and sales did not change much from the beginning to the end of the experiment, and the format of the notes did not change either. Overall, the number of characters and the format in the notes roughly converged for both pairs and dialogue tasks. See Fig. 5(a) and (c) for the format of the notes at the end of the experiment when they seem to have converged.

4.3. Analysis of the Notes

Since we did not specify how the notes were to be written, the operators used not only text but also graphical elements such as delimiter lines, circles, and arrows, as seen in Fig. 5(a) and (c). The arrows were typically used to represent relationships between utterances, such as questions and answers. This format is

similar to that of adjacency pairs (Schegloff and Sacks, 1973) with some extension. The conventional adjacency pairs consist of two adjacent turns of different speakers, and the first pair part triggers the second pair part (e.g., question-answer). The adjacency pairs have been expanded to include a case where the first pair part of a two-turn exchange is a weak trigger for the second pair part, such as when a comment is followed by its response (Schegloff, 2007; Stivers, 2013). According to Tsui (1989) and Enomoto et al. (2005), the adjacency pairs can be expanded to cope with pre/post-expansions and third pair parts (e.g., initiation-response-follow-up). The notes here included these types of adjacency pairs.

Most of the operator and user utterances in the notes were separated into left and right and recorded in the form of adjacency pairs. We found 158 adjacency pairs in the notes and manually classified the instances into conventional adjacency pairs (50 instances), extended adjacency pairs containing comment and follow-up (59 instances), and extended adjacency pairs containing three or more parts (49 instances). We had quite a few adjacency pairs containing the operator's comment or follow-up in the notes; such information seemed useful for the operators to improve the consistency of their own utterance content.

An example note for the consultation dialogue task is shown in Fig. 5(b). This note is more straightforward than the one in the chat task because consultation is a task-oriented topic, which means the intention of the speakers is clear and the conversation often centers on simple questions and answers. In the lower half of the note, we can see the information is written in the format of questions and answers (i.e., adjacent pairs), as in the case of chat, but in the upper half, the information is written in the form of key-value pairs (similar to slots in information states (Larsson and Traum, 2000)), e.g., "Length of stay | Four days and three nights". This format also appeared in the notes for sales (e.g., Fig. 5(c)). In consultation and sales tasks, the user information that the operator needs to obtain is often predetermined. Therefore, in addition to the adjacent pair format, the key-value pair was considered effective in these tasks because it allows an operator to see at a glance what information the operators have already gleaned and what information they have not yet gleaned.

The notes shown in Fig. 5 are from the end of the experiment, when the number of characters had converged. Figure 6 shows the notes of dialogues at the beginning of the experiment for comparison. The notes in the early part of the experiment seem rather unstructured. However, as the experiment progressed, the notes became more structured—e.g., utterances were separated into left and right parts, utterances were connected as adjacent pairs, and independent key-value pairs were created—indicating that the notes were made in a format suitable for dialogue handover.

O1	Um, well, we're going to go through the airport, um, the city, um, roughly so, um, about two or three hours, about three hours, I guess.
U	Hmm.
O1	Uh-huh.
U	I see, so it's not impossible to visit everything in three days, is it?
O1	Yes, that's true.
O1	Uh-huh.
O1	You're right.
–	<switching>
O2	Oh, sorry, I'm getting a bit of a signal.
U	Hmmm.
O2	Once again, please.
U	I think I'd like to stay in the town of Kobarid.
O2	Kobarid.

(a) Example of a dialogue in which the operators could not be switched smoothly. According to the interview, the operator did not know what to say even though he looked at his note, so he pretended that he could not hear the user's speech due to a poor signal and asked the user to speak again.

U	Uh, uh, the name of its maker is Gibson.
O1	Oh, it's a really expensive one, yes.
U	I got excited and bought it.
O1	Ah, yes, it's a bit pricey, isn't it?
U	That's right.
O1	I see.
O1	Heh, Gibson's, what model is it?
–	<switching>
U	Well, I don't remember the model number, but it's a black guitar.
O2	Black!
U	That's right.
O2	Gibson.
O2	Heh.

(b) Example of a dialogue in which the operators were smoothly switched. The switched operator (O2) was able to understand that the previous operator was talking about Gibson guitars.

Figure 7: Example of dialogue immediately before and after switching. Operator before the switch is indicated by O1, operator after the switch by O2, and user by U. The above dialogues have been translated from the original Japanese to English by the authors.

4.4. Analysis of Discussions and Interviews

During the discussions, one operator mentioned that after switching, she referred to only the last line of the note (which seemed to be the most recent information) because she did not immediately have time to read the whole note carefully. Then, after uttering something else, from the next turn onward, she mentioned that she went back through to grasp the history of the note. This indicates that it is necessary for the operator to have two types of information in order to immediately respond and to understand the dialogue history. One of the operators parroted back the most recent utterance of the user immediately after the switch. Parroting like this

Word 4-grams	Translation
ですよねー。 / そうですよねー / そうですねー、	I see. Yeah, yeah.
えええええ	Ummm.
ちょっと今電波が / ごめんなさいちょっと今電波	Sorry, I think we have a bad connection.
あ、 ごめんなさい、 / あ、 ごめんなさいちょっと	Oh, sorry.
ごめんなさい、もしもし。	Sorry, can you hear me?

Table 4: Word 4-grams whose frequency of occurrence immediately after switching was significantly higher in the operator's utterance.

is helpful as a first utterance after switching because it allows the operator to confirm what the user is saying and possibly elicit additional information.

During the interview, one operator stated that he was careful not to make mistakes such as contradiction or unintentional repetition of utterances that had already been made. However, there were occasions when he repeated the same questions. Inconsistent utterances were also frequent in both pairs of operators. For example, during a chat about chocolate, the operator before the switch said, "I think it's delicious, too", but the operator after the switch made a contradictory self-disclosure, saying, "I'm not a fan of sweets". This operator stated that it was difficult for two operators to behave as one unless they knew the personality of the other. This also highlights the importance of consistency in the dialogue between operators.

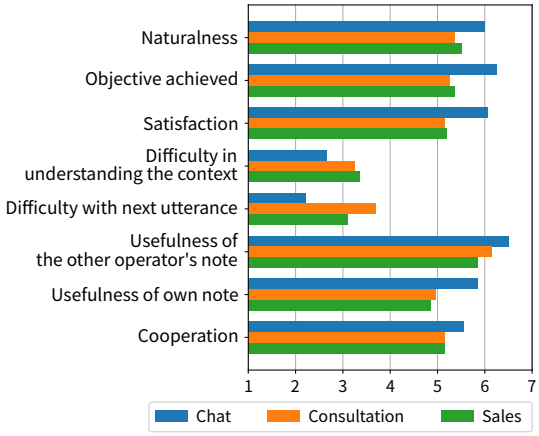
4.5. Analysis of Utterances After Switching

Figure 7 shows an example of dialogue before and after switching. It seems there were characteristic words in the vicinity of the operator switching. Therefore, we investigated what kind of expressions appeared there by mining word 4-grams.

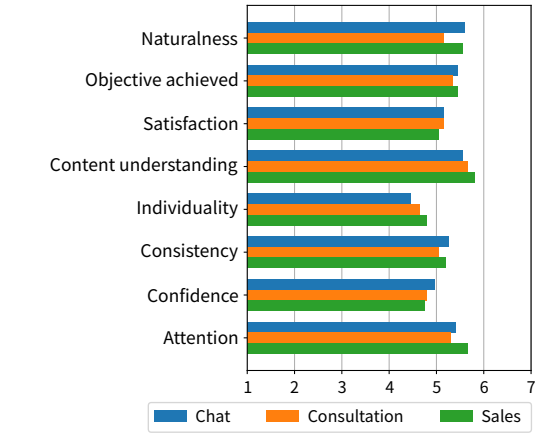
We compared the frequency of the word 4-gram between the operator's utterances made within ten seconds of the operator switching and those made at other times. For word segmentation, we used the morphological analyzer MeCab[5] with the IPAdic dictionary.

We created 2×2 tables that show two groups of operator utterances, immediately spoken after the operator switching or in another timing, and whether a certain word 4-gram appeared or not. To determine whether a certain word 4-gram was more likely to appear immediately after the switching, we conducted a Fisher's exact probability test. Table 4 lists the word 4-grams that appeared at a significantly higher frequency between after the switching and the other timing at the 1% level. As we can see, the word 4-grams that appeared frequently in the utterances immediately after the switching showed an increase in the number of back-channels (e.g., "I see." and "Yeah, yeah."), suggesting that the operator was prompting the user to continue the utterance. In addition, there were phrases such as "I'm sorry, I can just barely hear you because you're breaking up", indicating that the user's speech could not be heard. This result demonstrates that it is difficult to think of the following utterance immediately after the handover, highlighting the importance of having the optimal information to support the handover.

[5]https://taku910.github.io/mecab/



(a) Operator answers

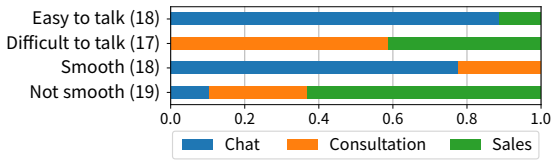


(b) User answers

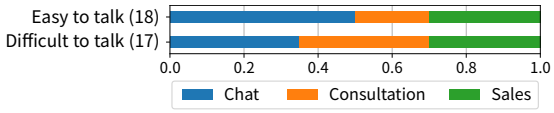
Figure 8: Averaged questionnaire scores. The questionnaire items were on a 7-point Likert scale.

4.6. Analysis of Questionnaires

Figure 8 shows the results of the questionnaire on dialogue. We can see that, especially in the chat task, operators had less difficulty in understanding the context and making the next utterance. We conducted Wilcoxon rank sum tests with Bonferroni correction and found a significant difference at the 1% level between chat and consultation in terms of the difficulty with the next utterance, and between chat and sales as to the usefulness of the other operator's note. It seems that adjacent pairs, which were frequently used in the notes, were particularly effective in handling chat. In the case of the consultation and sales tasks, the usefulness of the note written by the other operator was as high as that of chat, at around six points out of seven.



(a) Operator answers



(b) User answers

Figure 9: Proportion of dialogue tasks operators and users preferred. The number of valid answers is shown in parentheses.

The evaluations from the users were typically around four out of seven points for all three types of dialogue, indicating that they were able to interact with the operator reasonably well even when the operators periodically switched.

We investigated the correlation between the questionnaire on dialogue and the number of characters in a note. For the operator's answers, a positive correlation was found between the number of characters in a note and the evaluation of the usefulness of the other operator's notes ($r = 0.36$). For the user's answers, positive correlations were found between the number of note characters and individuality ($r = 0.38$), consistency ($r = 0.35$), and confidence ($r = 0.42$). These results suggest that notes with a large number of characters include helpful information for an operator to speak in a confident manner and were thus useful, enabling operators to produce speech with a sense of individuality, maintain consistency in speech content, and give users a sense of trust.

Figure 9 shows the results of the questionnaire administered after each of the three dialogue tasks, with 20 answers each from operators and users indicating which of the three tasks they felt best matched a given questionnaire statement. Since the questionnaire was descriptive (not multiple choice), we had answers such as "multiple dialogue tasks are applicable" or "nothing in particular". We excluded these answers from our analysis and used only those including just one dialogue task name. The figure shows the rate of valid answers in which each dialogue task was specified. Among the three types, the operators answered that chat was the easiest for talking and that cooperation among the operators was smoothest. Consultation was the most difficult for talking, while sales was the least smooth in cooperation. Although the percentage was lower than that of the operators, about half of the users also answered that chat was easiest for talking. These results suggest that the information necessary to take over a dialogue was successfully included in the notes for chat, which consisted primarily of adjacency pairs. Users answered that the degree of difficulty was roughly the same for all dialogue tasks, suggesting that the notes for consultation and sales, which included both adjacency pairs and key-value pairs, were also useful. However, the operators felt that handovers of consultation and sales dialogues were not smooth. It was implied that these tasks, which require task completion, are relatively difficult to hand over.

5. Conclusion

In this paper, we experimentally investigated the information needed for handing over a dialogue in a situation where two operators switched with each other and exchanged a note. We conducted handover experiments for chat dialogue, consultation dialogue, and sales dialogue and analyzed the notes made by the operators.

For handovers, the operators need to track basic exchanges (who said what to whom) and avoid inconsistency of their own utterances. To this end, adjacency pairs with their associated speakers, which can express basic exchanges and include the comments and follow-up utterances of speakers, were found to be useful among the various representations of dialogue. In task-oriented dialogues, such as consultation and sales, key-value pairs, which are similar to slot information in information states, were also found to be useful because they make it possible to grasp at a glance the information that has already been mentioned in tasks. Topic segments and discourse segments were not observable in the notes in our experiment; although they provide structure and give a global view of the entire dialogue, such information may not be needed to facilitate the handover of an ongoing dialogue at hand. It may also be possible that the operators were unable to create elaborate structures while continuing the dialogue because it was too cognitively demanding.

As future work, we aim to implement these findings in an interface suitable for dialogue handover. To this end, several underlying techniques will be necessary, such as extracting adjacency pairs from the dialogue history and extracting certain information in the form of key-value pairs from the dialogue history, as in the dialogue state tracking used in task-oriented dialogue systems (Williams and Young, 2007; Williams et al., 2014). It may also be necessary to format the extracted adjacency pairs in a more concise way for understanding the content quickly. Techniques such as identifying the dialogue acts of utterances in adjacency pairs, summarizing the dialogue for each adjacency pair, and including key-value pairs in the summary could all be utilized. Once the interface has been created, we plan to evaluate its usefulness with human participants in real handover situations, such as immediately after the dialogue breakdown (Higashinaka et al., 2016) by the system.

6. Acknowledgements

This work was supported by JST Moonshot R&D Grant number JPMJMS2011.

7. Bibliographical References

Adiwardana, D., Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., et al. (2020). Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Arimoto, T., Higashinaka, R., Tanaka, K., Kawanishi, T., Sugiyama, H., Sawada, H., and Ishiguro, H. (2020). Collection and analysis of dialogues provided by two speakers acting as one. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 323–328.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In *Proceedings of Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901.

Enomoto, M., Den, Y., and Matsusaka, Y. (2005). Annotation of dialog acts and addressees in three-party conversations and its preliminary analysis. *JSAI Technical Report, Special Interest Group on Spoken Language Understanding and Dialogue Processing*, 45:7–12. (In Japanese).

Gorin, A. L., Riccardi, G., and Wright, J. H. (1997). How may I help you? *Speech communication*, 23(1–2):113–127.

Grosz, B. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.

Hearst, M. A. (1997). Text tiling: Segmenting text into multi-paragraph subtopic passages. *Computational linguistics*, 23(1):33–64.

Higashinaka, R., Funakoshi, K., Kobayashi, Y., and Inaba, M. (2016). The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pages 3146–3150.

Huang, T.-H., Chang, J. C., and Bigham, J. P. (2018). Evorus: A crowd-powered conversational assistant built to automate itself over time. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Jiang, F., Fan, Y., Chu, X., Li, P., Zhu, Q., and Kong, F. (2021). Hierarchical macro discourse parsing based on topic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13152–13160.

Kawahara, T., Muramatsu, N., Yamamoto, K., Lala, D., and Inoue, K. (2021). Semi-autonomous avatar enabling unconstrained parallel conversations–seamless hybrid of WOZ and autonomous dialogue systems–. *Advanced Robotics*, pages 1–7.

Larsson, S. and Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, 6(3-4):323–340.

Ou, W. and Lin, Y. (2020). A clarifying question selection system from NTES_ALONG in Convai3 challenge. *arXiv preprint arXiv:2010.14202*.

Pichl, J., Marek, P., Konrád, J., Lorenc, P., Ta, V. D., and Šedivỳ, J. (2020). Alquist 3.0: Alexa prize bot using conversational knowledge graph. *arXiv preprint arXiv:2011.03261*.

Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Smith, E. M., Boureau, Y.-L., and Weston, J. (2021). Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325.

Schegloff, E. A. and Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4):289–327.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*, volume 1. Cambridge university press.

Stivers, T. (2013). Sequence organization. In Jack Sidnell et al., editors, *The handbook of conversation analysis*, chapter 10, pages 191–209. Blackwell Publishing Ltd.

Tsui, A. B. (1989). Beyond the adjacency pair. *Language in Society*, 18(4):545–564.

Walker, M., Langkilde, I., Wright, J., Gorin, A. L., and Litman, D. (2000). Learning to predict problematic situations in a spoken dialogue system: experiments with How may I help you? In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 210––217.

Williams, J. D. and Young, S. (2007). Partially observable Markov decision processes for spoken dialog systems. *Computer Speech & Language*, 21(2):393–422.

Williams, J. D., Henderson, M., Raux, A., Thomson, B., Black, A., and Ramachandran, D. (2014). The dialog state tracking challenge series. *AI Magazine*, 35(4):121–124.

Yamashita, S. and Higashinaka, R. (2021). Analysis of summaries that enable a smooth intervention in chat-oriented dialogue. *JSAI Technical Report, Special Interest Group on Spoken Language Understanding and Dialogue Processing*, 93:86–91. (In Japanese).