Collection and Analysis of Dialogues Provided by Two Speakers Acting as One

Tsunehiro Arimoto[1], **Ryuichiro Higashinaka**[1], **Kou Tanaka**[1], **Takahito Kawanishi**[1]
Hiroaki Sugiyama[1], **Hiroshi Sawada**[1], **and Hiroshi Ishiguro**[2]
[1]NTT Communication Science Laboratories
[2]Osaka University
{arimoto.tsunehiro.ub, ryuichiro.higashinaka.tp, kou.tanaka.ef
takahito.kawanishi.fx, hiroaki.sugiyama.kf, hiroshi.sawada.wn } @hco.ntt.co.jp
ishiguro@irl.sys.es.osaka-u.ac.jp

Abstract

We are studying a cooperation style where multiple speakers can provide both advanced dialogue services and operator education. We focus on a style in which two operators interact with a user by pretending to be a single operator. For two operators to effectively act as one, each must adjust his/her conversational content and timing to the other. In the process, we expect each operator to experience the conversational content of his/her partner as if it were his/her own, creating efficient and effective learning of the other's skill. We analyzed this educational effect and examined whether dialogue services can be successfully provided by collecting travel guidance dialogue data from operators who give travel information to users. In this paper, we report our preliminary results on dialogue content and user satisfaction of operators and users.

1 Introduction

Such dialogue services as counseling (Dowling and Rickwood, 2013) are often provided through telecommunication systems that enable speakers (typically called operators) to talk from remote places (Crabtree et al., 2006; Sakamoto et al., 2007; Yamashita et al., 2011; Kristoffersson et al., 2013). For such services to be more productive, it is desirable that the skills of the operators are improved.

In this paper, we propose a unique learning style in which multiple operators with different skills cooperate and pretend to be one person (Fig. 1). For two operators to effectively act as one, each must adjust his/her conversational content and timing to the other. In this style, each operator may experience the conversational content of his/her partner as if it were his/her own, creating efficient and effective learning of the other's skill. Users also benefit; they do not have to interact with a lot of operators and can establish one-to-one relationships. There were studies that aimed at increasing the perceived number of speakers for better interaction despite that there is only a single operator (Yamane et al., 2011; Arimoto et al., 2014); our idea here is the opposite.

ID	Spk	Utterance
1	**U**	**Hello. I am planning trips to Nara and Osaka prefectures. What sightseeing spots do you recommend?**
2	GN	Hello. In the Nara area, I recommend Todaiji Temple and Nara Park.
3	**U**	**I see. How can I get to them?**
4	GN	You can walk to Todaiji Temple from Kintetsu Nara Station through Nara Park.
5	**U**	**Thank you. How about Osaka?**
6	GN	(Your turn.)
7	GO	(Ok.)
8	GO	Well, in Osaka, I recommend Osaka Castle and Universal Studios.
9	**U**	**Those are both famous.**
10	GO	You can easily get to them by train.
11	**U**	**I'm glad they are so convenient. By the way, in Nara, do you recommend any restaurants where I can eat local food around those two spots?**
12	GO	(Why don't you answer?)
13	GN	(Sure.)
14	GN	I recommend Asuka Nabe.
15	**U**	**I see. Any idea how much it costs?**

Figure 1: Example of Mixto1 condition where two guides with different skills pretend to be one guide who talks to a user (U). One guide has knowledge about travel in Nara (GN), and the other knows Osaka (GO). For readability, user utterances are shown in bold. Parentheses represent invisible to a user.

Many prior studies exist where multiple actors work together to provide dialogue services. Cooperative architectures with multiple agents or human operators have attracted attention with regards to the development of dialogue systems (Lin et al., 1999; Komatani et al., 2009;

, pages 290 - 295

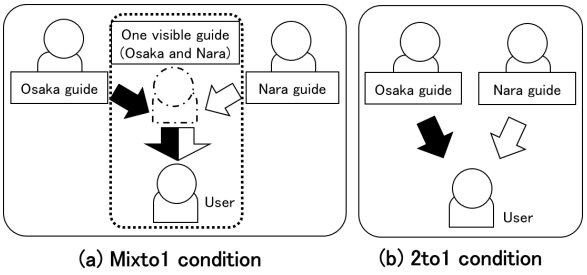

Figure 2: Cooperation style of Osaka and Nara guides under Mixto1 and 2to1 conditions

Nakano et al., 2011) as well as Wizard-of-Oz systems (Marge et al., 2016; Abbas et al., 2020). Users talking with a dialogue assistant controlled by multiple speakers on the cloud are reported to receive more reasonable responses (Lasecki et al., 2013). However, no research has examined the basic effect of behaving as one speaker on the satisfaction of the operators and their interlocutors. It remains especially unclear whether multiple operators who are acting as one promote mutual skill learning.

The following is the contribution of this study. First, we show a method for collecting text-chat dialogues in which two speakers acting as one person. Second, we show the basic effects of two speakers who are pretending to be just one person on the dialogue's content and the satisfaction of the operators and the interlocutors.

2 Collection of text chats in which two speakers act as one

2.1 Dialogue design

Our study focuses on the dialogue services of two human operators with different knowledge. With different knowledge, the two operators can provide a larger variety of information than when they are separate. We collected travel guide text-chat dialogues about two neighboring prefectures. The dialogues were conducted by either one or two operators. We categorized the travel guidance knowledge for one prefecture as each operator's skill. We have the following three conditions for conducting a dialogue:

Mixto1 condition Two operators with different specialties (as their skills) acted as one speaker. For example, we paired an operator who is familiar with Osaka prefecture and another who is familiar with Nara prefecture. Nara and Osaka are geographically adjacent. They acted as one visible guide with knowledge of both prefectures (Fig. 2(a)).

2to1 condition Two operators with different specialties took turns talking directly (Fig. 2(b)) with one user in a three-party dialogue. This condition was collected as a baseline to evaluate the validity of the Mixto1 condition.

1to1 condition One operator gave recommendation to one user about two prefectures. The operator has much knowledge about one of them, but the other is outside his/her skill set.

Collaborative dialogues (Mixto1 and 2to1 conditions) are expected to positively affect the operators' learning. We collected the 1to1 condition dialogues before and after the Mixto1 and 2to1 conditions to examine such educational effects.

2.2 Environment

All the speakers used Slack[1] to communicate in a text-chat format. They played either a guide (operator) or a user.

In the Mixto1 condition, two guides acted as one guide and interacted with one user. Each guide opened two Slack windows in one display. One window was used to interact with the user, and the other was used to consult with the other guide. The guides discussed their strategy for talking with the user in a window hidden from the user. The user opened a window to interact with the guide in one display and talked with both guides about his/her trip to the two pre-designated prefectures. The two guides used the same account to talk to the user; the user didn't realize he/she was talking to two guides.

In the 2to1 condition, two guides and one user also participated in the dialogue as in the Mixto1 condition. However, both talked to the user using different accounts. Each guide opened a window to interact with the user without opening an additional window to just interact with the other guide.

In the 1to1 condition, one operator and one user each opened a window and directly interacted with each other.

2.3 Subjective questionnaires

Since it is unclear how our collected interactions affected the satisfaction of the guides, they answered a 12-item subjective questionnaire to assess task achievement and their impressions of

[1]https://slack.com

Figure 3: Data collection. Numbers in parentheses represent the number of interactions.

each conversation. For the Mixto1 and 2to1 conditions, the guides also answered three items about their impressions of performing the conversation as one or two people. They also freely described their experience at the end of Session 2 (Section 2.4). The users answered ten items regarding their impressions of the task achievement and the conversations.

2.4 Data collection

We recruited speakers to act as operators or users. The operators and users were paid for their participation. All dialogues were conducted in Japanese. Sixteen operators participated as guides. Operators were assigned to their home prefecture as their specialty (we assume that operators were knowledgeable about their home prefectures). Their ages ranged from 20 to 50 years, with six males and ten females. Two guides of the same gender from neighboring prefectures were paired.

Forty-eight speakers (16 males and 32 females) whose ages ranged from 20 to 50 participated in the dialogues as users. Each participated in a travel guide dialogue outside their home prefecture.

We collected the data over three sessions (Fig. 3). All the guides participated in all three sessions. Sixteen were divided into two groups of eight; the M group having the Mixto1 condition and the T group having the 2to1 condition in Session 2. Users participated in only one of the sessions and talked three times with different guides or guide pairs under the same condition. Each dialogue lasted ten minutes. We collected 144 travel text-chat dialogues and questionnaires from each guide and each user. The following are the descriptions of Sessions 1 to 3:

Sessions 1 and 3 All guides talked under the 1to1 condition. Each guide had text chats three times with a different user in each dialogue. We collected 48 dialogues for each session.

Session 2 The M group's guide pair worked under the Mixto1 condition and the T group worked under the 2to1 condition. Each guide pair had text chats six times with a different user in each dialogue. Therefore, we collected 24 Mixto1 dialogues and 24 2to1 dialogues.

3 Analysis

3.1 Approach

Evaluation of dialogue flows Using the collected text chat, we qualitatively analyzed how the guides facilitated the travel decisions under each condition. Under the 1to1 condition, the guides had limited knowledge that assisted them with travel to prefectures outside their specialty. Under the Mixto1 and 2to1 conditions, the operator had the opportunity to provide trip guidance while talking in turns with the other guide. We observed how the guides made recommendations based on the conditions.

Number of guide utterances for non-specialty prefectures The guides touched on the expertise of the other guides under the Mixto1 and 2to1 conditions. These guides may have gained information about the non-specialized prefectures from the conversations of the other guides, educating them about these unfamiliar prefectures. We analyzed whether the Mixto1 condition, acting as a single guide, increases the utterances of the non-specialized prefectures of guides.

3.2 Results

Dialogue flows The actual examples of collected dialogues for the Mixto1 and 2to1 conditions are shown in Figs 1 and 4 (translated from Japanese to English by authors).

In the 1to1 condition, the guide talked one-to-one with one user. In some scenes, the guide was unable to answer questions outside his specialty. For example, the guides frequently said "I'm sorry I don't know" in the conversation.

In the Mixto1 example (Fig 1), two guides provided travel recommendations for Nara and Osaka prefectures. For "I am planning trips to Nara and Osaka prefectures. What sightseeing spots do you recommend? (ID = 1)," the Nara guide introduced Nara ("Hello. In the Nara area, I recommend Todaiji (ID = 2)") and the Osaka guide introduced Osaka ("Well, in Osaka, I recommend Osaka Castle and Universal Studios (ID = 8)"). By using the window that was hidden from the user, the guides could consult when to switch among themselves (e.g., "Your turn (ID = 6)" and "Ok (ID = 7)").

ID	Spk	Utterance
1	**U**	**Nice to meet you. I am thinking of traveling around Fukuoka and Kumamoto for 3 or 4 nights. I'd like to go to Aso Nature Land, Dazifu Tenmangu, and the food stalls in Nakasu. What other places do you recommend?**
2	GK	If you have time, I recommend Kurokawa Onsen.
3	**U**	**I see. I also want to go to a hot spring. This'll be my first visit to Fukuoka.**
4	GF	Hello. If you have time, how about Moji Port in Kitakyushu City because its retro streets are cute. Lots of fancy cafes and souvenir shops, too.
5	**U**	**Oh, that sounds nice. I love eating, too.**
6	GF	If you are looking for gourmet food, the iron-pot gyoza and mizutaki around Haruyoshi are delicious in Fukuoka.
7	GK	In Kumamoto, the Kumamoto ramen, basashi, and red ox dishes are famous.

Figure 4: Example of 2to1 condition: U, GF, and GK represent user, guide for Fukuoka, and guide for Kumamoto. For readability, user utterances are shown in bold.

In the 2to1 condition, the two guides talked individually to directly help the user. Figure 4 shows the travel guide dialogue for Kumamoto and Fukuoka prefectures by the Kumamoto and Fukuoka guides. Both guides talked about their specialty. The Fukuoka guide said, "the iron-pot gyoza and mizutaki around Haruyoshi are delicious in Fukuoka (ID = 6)." The Kumamoto guide said, "In Kumamoto, the Kumamoto ramen, basashi, and red ox dishes are famous (ID = 7)."

These observations show that the guides had the opportunity to provide trip assistance while speaking in turns with the other guide under the Mixto1 and 2to1 conditions.

Number of guide utterances for non-specialty prefectures We annotated whether each utterance in the dialogue was related to each of the two prefectures and counted the number of utterances of the guides for their non-specialized prefectures. For each group (M and T), we analyzed whether there was a difference in the number of utterances in Sessions 1 and 3 before and after completing Session 2.

A Wilcoxon's rank-sum test showed that the M group under the Mixto1 condition showed a significant tendency to increase the number of utterances regarding non-specialized prefectures 1to1 of Session 1 (M group) = 2.5, 1to1 of Session 3 (M group) = 4.0, W = 198, $p < .1$). On the other hand, we found no significant difference in the T group who experienced the 2to1 condition（1to1 of Session 1 (T group) = 1.5, 1to1 of Session 3 (T group) = 3.0, W = 227, p = n.s.). This result suggests that the M group guides gained knowledge about their non-specialties by experiencing the Mixto1 conditions.

4 Subjective Impressions of Speakers

We analyzed the overall satisfaction impressions of the guides and users on a 7-point Likert scale (7 = totally agree, 1 = totally disagree).

4.1 Approach

Guide satisfaction Our study focused on the situation where two guides talk as one. Such a situation might be confusing for guides and users. To ensure that the guides did not have any difficulty speaking under this condition, we used the following statement: "When I talked to the user, I sometimes felt it was difficult."

In the Mixto1 condition, two guides talked as one. By sharing the dialogue context as one operator, each operator may experience the conversational content of his/her partner as if it were his/her own, creating efficient and effective learning of the other's skill. In the Mixto1 condition, the guides may also be more aware of cooperating with the other guides and deepen their mutual trust.

We used the following three items to evaluate the guide's satisfaction with the other guide's cooperation: Statement (a) assessed feelings of respect for the other guide: "I felt a sense of trust in the other guide." To evaluate the ease of cooperation with the other guide, we used statement (b): "I was able to work with the other guide." To evaluate the impressions of learning from the other guide, we used statement (c): "I learned from the other guide's responses."

User satisfaction The easy-to-talk impressions felt by users under the Mixto1 and 2to1 conditions may differ. In the latter, the user distinguishes between the two guides and interacts in a multi-party manner. However, the user does not distinguish between them in the Mixto1 condition. This difference might affect the user's speaking ease. To evaluate whether users felt it was difficult to talk, we used questionnaire item (d): "There were times when I felt it was hard to talk."

We also evaluated whether users felt they accomplished their task with questionnaire item (e):

"Through the dialogue with the guide(s), I obtained useful information" to evaluate whether the users obtained the necessary knowledge for their travel.

4.2 Results

Guide satisfaction To analyze the impressions of the guides' difficulty in speaking, we calculated the median of each condition. The median of each condition was lower than four points. This indicates that the guides did not perceive particular difficulty in speaking.

For their impressions of cooperating with another guide, we compared (a), the trust of another guide, under the Mixto1 and 2to1 conditions. Wilcoxon's rank-sum test showed that the Mixto1 condition was significantly higher than the 2to1 condition (Mixto1 = 6, 2to1 = 5, W = 1520.5, $p<.05$).

We also compared (b), measure of cooperation satisfaction, with the Mixto1 and 2to1 conditions. The Mixto1 condition was significantly higher than the 2to1 condition (Mixto1 = 6, 2to1 = 4, W = 1831, $p<.05$).

The Mixto1 and 2to1 conditions were also compared for (c), an evaluation item of learning impression. The Mixto1 condition was significantly higher than the 2to1 condition (Mixto1 = 6, 2to1 = 5, W = 1445, $p<.05$).

From the above results, the guides' satisfaction was higher in the Mixto1 condition than in the 2to1 condition. The guides felt a sense of cooperation and trust with the other guide, adding that under the Mixto1 condition, they acquired more knowledge than under the 2to1 condition.

One possible factor that resulted in such positive impressions for the Mixto1 condition was that the guides were engaged in first-person conversations. Probably they quickly became absorbed in the conversations because the users acted like just one guide. Perhaps the guides felt that they had acquired knowledge because it was easy to regard the utterances of the other guides as their own. In the future, we must clarify which factor deepens the guides' impressions of subjective learning by scrutinizing the dialogue content.

In addition, it may also be necessary to examine the effect of a hidden channel used by the guides because it may have had particular effects on the cooperation of the guides.

User satisfaction We did not find a significant difference in (d), the users' perceived difficulty of speaking, in a Wilcoxon's rank-sum test that compared the Mixto1 and 2to1 conditions (Mixto1 = 2, 2to1 = 3, W = 235, p = n.s.). Both median values were lower than four (= neither), suggesting that they did not find it difficult to talk under either condition.

Next we analyzed (e), the impression of the users' information collection. When the Mixto1 and 2to1 conditions were compared, no significant difference was detected (Mixto1 = 6, 2to1 = 6, W = 258, p = n.s.). Both conditions had high scores. Perhaps the task of acquiring travel knowledge was relatively easy. Differences might surface in more difficult tasks.

In this experiment, we identified no significant differences in the user satisfaction between the Mixto1 and 2to1 conditions. However, we also found no evidence that the Mixto1 condition negatively impacted the users. Whether Mixto1 can improved the dialogue quality must be investigated with another situation in the future.

5 Conclusion

We evaluated a situation in which two operators with different skills acted as one. We collected travel guide dialogues where two operators acting as one speaker, as two speakers, and alone. We evaluated the contents under each condition as well as the satisfaction of the operators and users. The operators experienced increased satisfaction with their learning and cooperation. The users were not dissatisfied with the situation of two operators speaking as one. It is suggested that the proposed cooperation style gives operators an opportunity to engage in advanced dialogue services as well as to learn the skills of the other operators.

In the future, we must scrutinize how the operators increased their satisfaction with learning and evaluate what kind of knowledge sharing occurred between the operators. We also need to examine a combination of other kinds of skills.

Acknowledgments

This work was supported by JST-Mirai Program Grant Number JPMJMI18C6, Japan.

References

Tahir Abbas, Vassilis-Javed Khan, and Panos Markopoulos. 2020. Coz: A crowd-powered system for social robotics. *SoftwareX*, 11:100421.

Tsunehiro Arimoto, Yuichiro Yoshikawa, and Hiroshi Ishiguro. 2014. Nodding responses by collective proxy robots for enhancing social telepresence. In *Proceedings of the Second International Conference on Human-Agent Interaction*, pages 97–102.

Andy Crabtree, Jacki O'Neill, Peter Tolmie, Stefania Castellani, Tommaso Colombino, and Antonietta Grasso. 2006. The practical indispensability of articulation work to immediate and remote helpgiving. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 219–228.

Mitchell Dowling and Debra Rickwood. 2013. Online counseling and therapy for mental health problems: A systematic review of individual synchronous interventions using chat. *Journal of Technology in Human Services*, 31(1):1–21.

Kazunori Komatani, Naoyuki Kanda, Mikio Nakano, Kazuhiro Nakadai, Hiroshi Tsujino, Tetsuya Ogata, and Hiroshi G Okuno. 2009. Multi-domain spoken dialogue system with extensibility and robustness against speech recognition errors. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 9–17.

Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, 2013.

Walter S Lasecki, Rachel Wesley, Jeffrey Nichols, Anand Kulkarni, James F Allen, and Jeffrey P Bigham. 2013. Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 151–162.

Bor-shen Lin, Hsin-min Wang, and Lin-shan Lee. 1999. A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history. In *Proceedings of the 1999 IEEE Workshop on Automatic Speech Recognition and Understanding Workshop (ASRU)*.

Matthew Marge, Claire Bonial, Kimberly A Pollard, Ron Artstein, Brendan Byrne, Susan G Hill, Clare Voss, and David Traum. 2016. Assessing agreement in human-robot dialogue strategies: A tale of two wizards. In *Proeedings of Intelligent Virtual Agents*, pages 484–488.

Mikio Nakano, Shun Sato, Kazunori Komatani, Kyoko Matsuyama, Kotaro Funakoshi, and Hiroshi G Okuno. 2011. A two-stage domain selection framework for extensible multi-domain spoken dialogue systems. In *Proceedings of the SIGDIAL 2011 Conference*, pages 18–29.

Daisuke Sakamoto, Takayuki Kanda, Tetsuo Ono, Hiroshi Ishiguro, and Norihiro Hagita. 2007. Android as a telecommunication medium with a human-like presence. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 193–200.

Masaru Yamane, Michiya Yamamoto, and Tomio Watanabe. 2011. Development of a real-space sharing edutainment system based on communication support with make-believe play. In *SICE Annual Conference 2011*, pages 2571–2574.

Naomi Yamashita, Hideaki Kuzuoka, Keiji Hirata, Shigemi Aoyagi, and Yoshinari Shirai. 2011. Supporting fluid tabletop collaboration across distances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2827–2836.