Speech and language resources for the development of dialogue systems and problems arising from their deployment

journal or publication title	Proceedings of the LREC 2018 Special Speech Sessions
page range	40-46
year	2018-05-09
URL	http://doi.org/10.15084/00001916

Speech and Language Resources for the Development of Dialogue Systems and Problems Arising from their Deployment

Ryuichiro Higashinaka[1], **Ryo Ishii**[1], **Narimune Matsumura**[1]
Tadashi Nunobiki[1], **Atsushi Itoh**[2], **Ryuichi Inagawa**[2], **Junji Tomita**[1]
[1]NTT Corporation, [2]NTT Data Corporation
higashinaka.ryuichiro@lab.ntt.co.jp, ishii.ryo@lab.ntt.co.jp, matsumura.narimune@lab.ntt.co.jp
nunobiki.tadashi@lab.ntt.co.jp, inagawar@nttdata.co.jp, itouats@nttdata.co.jp, tomita.junji@lab.ntt.co.jp

Abstract

This paper introduces the dialogue systems (chat-oriented and argumentative dialogue systems) we have been developing at NTT together with the speech and language resources we used for building them. We also describe our field trials for deploying dialogue systems on actual premises, i.e., shops and banks. We found that the primary problem with dialogue systems is timing, which led to our current focus on multi-modal processing. We describe our multi-modal corpus as well as our recent research on multi-modal processing.

Keywords: Chat-oriented dialogue system, argumentative dialogue system, deployment of dialogue systems, multi-modal processing

1. Introduction

We are seeing an emergence of dialogue systems in our daily lives. Many task-oriented dialogue systems, such as Siri, Cortana, and Alexa, have been in use in our daily lives, and there have been a number of non-task oriented ones for social and entertainment purposes (Onishi and Yoshimura, 2014; Vinyals and Le, 2015; Shang et al., 2016; Higashinaka et al., 2017a).

NTT has been working on dialogue systems for decades, and, in terms of research, we are now specifically focusing on chat-oriented dialogue systems. This is because chat is an important part in human-machine communication. According to the survey done by the National Institute for Japanese Language and Linguistics, more than 60% of our conversations can be classified as chat (Koiso et al., 2016). This means, if we do not equip dialogue systems with chat capability, they will not be able join our conversations most of the time, which makes it difficult for such systems to become our “partners”. In addition to the survey, it has also been pointed out that we tend to chat with systems, even though users are explicitly informed that the systems are task-oriented (Takeuchi et al., 2007). This means that, even for task-oriented dialogue systems, chat capability is necessary for them to be useful.

We first introduce our chat-oriented dialogue system that we are developing. Since the system has to handle open-domain utterances from users, it needs to have an abundance of knowledge, requiring a number of resources. We describe the speech and language resources we created to develop our chat-oriented dialogue system. In addition to our chat-oriented dialogue system, we describe our recent work on an argumentative dialogue system that can have discussions with people. The aim of creating this system is to investigate ways to make users more engaged in conversation; topics tend to transit from one to the other in chat, whereas discussion requires more attention on a certain discussion topic, making argumentation an ideal research subject. Second, apart from our research prototypes, we have also been conducting trials of dialogue systems with actual users, placing systems on premises, such as shops and banks. This paper presents two case studies of such trials. Finally, we describe our recent work on multi-modal processing because in our research and also from deployment experience, we found that timing is by far the key problem with current dialogue systems.

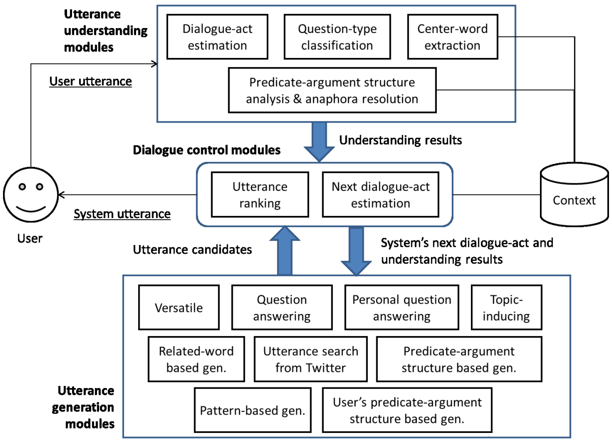
Utterance understanding modules
Dialogue-act estimation
Question-type classification
Center-word extraction
User utterance
Predicate-argument structure analysis & anaphora resolution
Understanding results
Dialogue control modules
Utterance ranking
Next dialogue-act estimation
Context
System utterance
User
Utterance candidates
System's next dialogue-act and understanding results
Versatile
Question answering
Personal question answering
Topic-inducing
Related-word based gen.
Utterance search from Twitter
Predicate-argument structure based gen.
Utterance generation modules
Pattern-based gen.
User's predicate-argument structure based gen.

Figure 1: System architecture of our chat-oriented dialogue system (see (Higashinaka et al., 2014) for details)

In Section 2, we describe our chat-oriented and argumentative dialogue systems. In Section 3, we describe our deployment of dialogue systems, covering two case studies. In Section 4, we describe our multi-modal corpus and our research regarding multi-modal processing. We summarize the paper and mention future work in Section 5.

2. Dialogue systems and resources

2.1. Chat-oriented dialogue system

Figure 1 shows the architecture of our chat-oriented dialogue system. The system has multiple modules, which can be classified into three blocks: utterance understanding, dialogue control, and utterance generation.

The system works in the following steps: given an input user utterance, utterance-understanding modules analyze the utterance, estimate a dialogue act and question type, extract center words (foci/topics in an utterance), and predicate argument structures (PASs). Dialogue-control modules receive the utterance-understanding results and determine the next dialogue act of the system. The utterance-understanding results and dialogue act of the system are fed to the utterance-generation modules to generate utterance candidates, which are finally ranked by the ranking module in the dialogue control. Finally, the top-rank utterance is selected to be output to the user (see (Higashinaka et al., 2014) for details of these modules).

Since we focus on open-domain conversation, we created a number of language resources for handling a variety of topics. For dialogue-act estimation, center-word extraction, and PAS analysis, we created training data to realize such functions with machine-learning methods. Specifically, on top of the chat dialogue data we collected, we carried out multiple annotations; namely, dialogue-act annotation, center-word annotation, and PAS annotation. We also carried out discourse relation annotation using the relations in the Penn Discourse Tree Bank (PDTB) (Miltsakaki et al., 2004). To generate a variety of system utterances, we created large-scale response rules in Artificial Intelligence Markup Language (AIML) (Wallace, 2009), which are used in the pattern-based generation module in Figure 1. We describe these resources below.

2.1.1. Chat dialogue corpus and its annotations

We use our chat-dialogue corpus as a base corpus. We collected 3,680 chat dialogues between two human users using a messenger interface. The total number of utterances is about 134K, and the number of users is 95. More than 1.2M words are included in the corpus. The length of a dialogue is about 36 utterances on average with about 9 words per utterance.

We sampled 20K utterances and carried out center-word annotation, in which noun phrases (NPs) denoting the foci/topics are annotated in utterances. We carried out dialogue-act annotation on all utterances in our chat-dialogue corpus. We used the dialogue-act taxonomy in (Meguro et al., 2010). The dialogue-act tag covers diverse utterances, making it suitable for open-domain conversation. There are 33 dialogue acts in the tag set. For PAS annotation, we sampled about 300 dialogues and annotated them with PASs; for each predicate, we mainly annotated `ga` (nominative), `wo` (accusative), and `ni` (dative) cases as well as several optional cases. We also carried out co-reference annotation, including zero-anaphora annotation (Imamura et al., 2014). Finally, for all utterances in the corpus, we carried out PDTB-style discourse-relation annotation. This chat-dialogue corpus is, as far as we know, by far the most well-annotated chat-dialogue corpus in Japanese. The annotations have been used to train models for center-word extraction, dialogue-act estimation, PAS extraction (including anaphora resolution), and discourse-relation detection. Discourse-relation detection has been found effective for ranking utterance candidates (Otsuka et al., 2017).

2.1.2. Large-scale response rules in AIML

We created large-scale response rules in AIML. We first created an initial rule set then revised it in the following manner. First, one text analyst created 149,300 rules by referring to our dialogue resources, mainly our chat-dialogue corpus described above. Then, an external judge subjectively evaluated the quality of the rules by inputting sampled utterances into a system loaded with the rules, and only when more than 90% of the responses were above average (over 6 points out of 10) was the rule-creation terminated. Then, we revised this rule set by using online evaluation where one external judge chatted for two turns with the system and evaluated the interactions subjectively. The rule-revision process terminated only when the judge was satisfied (same criterion as above) 90% of the time within 100 interactions. We ran eight iterations of this procedure to finalize the revised rule set. The entire revision process took approximately three months. At the end, the rule set contained 333,295 rules (categories in AIML) (see (Higashinaka et al., 2015) for details of this rule-creation process).

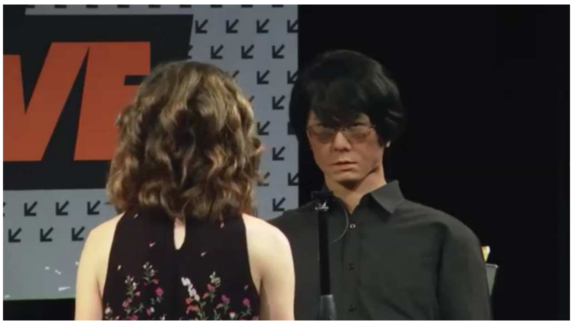

Figure 2: Geminoid HI-4 with our chat-oriented dialogue system at SXSW 2016. ©2015-2016 SXSW, LLC. This research was conducted in collaboration with Ishiguro laboratory of Osaka University.

2.1.3. Performances

We created two dialogue-system prototypes based on our architecture. One is Matsukoroid[1], which is an android robot that looks exactly like the famous TV personality Matsuko Deluxe in Japan. We incorporated our chat-oriented dialogue engine into the robot and let Matsuko Deluxe and his android chat with each other. This interaction was aired on Japanese television. The other is another android called Geminoid HI-4 (See Figure 2). We performed a live demonstration at South by South West (SXSW) in 2016. This system was an English port of our Japanese system; the overall architecture was the same with English data we newly created.

2.2. Argumentative dialogue system

Our chat-oriented dialogue system can maintain conversation by tracking center-words and by responding with large-scale rules as well as knowledge mined from the web. However, we also found that the content of a dialogue is rather superficial because the topics transit from one to the other, not going deeper into a topic. This sometimes made the dialogue less engaging for users.

[1] `https://naturaleight.co.jp/matsukoroid/`

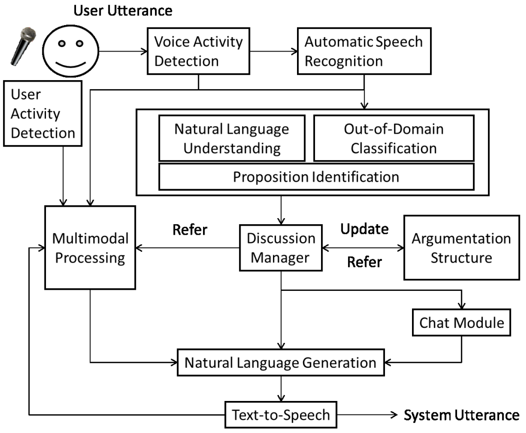

Figure 3: System architecture of our argumentative dialogue system (see (Higashinaka et al., 2017b) for details)

As our next step towards more engaging dialogue, we have been focusing on argumentative dialogue systems, in which users can engage in a discussion on a certain topic. Although much work has been done in argumentation mining (Lippi and Torroni, 2016), there has been less research on automated dialogue systems that can participate in discussion with users. We created large-scale "argumentation structures" as the knowledge of a system to conduct discussion. Our system uses such structures to generate supporting/non-supporting utterances as well as to keep track of the discussion.

Figure 3 shows the architecture of our argumentative dialogue system. At the core of the system is the argumentation structure, which is updated during the discussion, and from which the system's premises are generated.

2.2.1. Argumentation structures

We use a simplified version of the argumentation model described in (Gordon et al., 2007; Walton, 2013). The model has a graph structure, and nodes represent premises and edges represent support/non-support relationships between nodes. Each node has a natural language statement representing the content of its premise. We manually created several large-scale argumentation structures with each structure having more than 2,000 nodes. Each structure has two parts represented by main-issue nodes that enable the system to have opposing stances. Below the main-issue nodes, there are what we call viewpoints nodes that represent conversational topics. Under each viewpoint node, there are premise nodes that represent statements regarding each topic (see (Sakai et al., 2018) for details of our argumentation structures).

2.2.2. Performances

We integrated our argumentative system with an android and conducted a live demonstration at SXSW 2017 (See Figure 4). In our demonstration, two robots having opposite stances on a topic (e.g., which is the better living environment, east or west coast?) and three humans participated in a discussion[2]. Although there was some difficulty in controlling such multi-party conversation, since the argumentation structure was keeping track of the discussion and was updated appropriately on the basis of the utterances of the participants, we managed to conduct a reasonable demonstration.

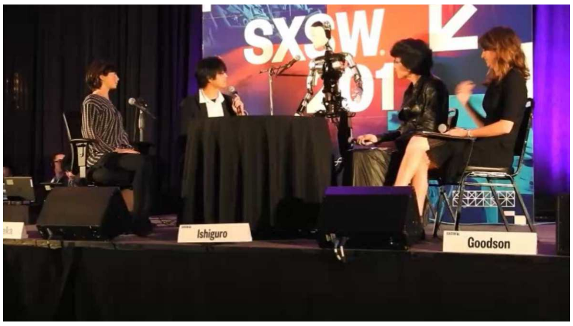

Figure 4: Two androids discussing with three humans at SXSW 2017. ©2016-2017 SXSW, LLC. This research was conducted in collaboration with Ishiguro laboratory of Osaka University.

2.3. Problems with our current systems

In our efforts in building chat-oriented and argumentative dialogue systems, we encountered the following difficulties.

- Since our systems are working on the text level, it was difficult to distinguish nuances in speech. For example, we expect question marks at the end of an utterance for a question in text, but it is not present in speech. Such para-linguistic information should be incorporated when considering the integration of text-based systems with androids that work on speech.
- We had difficulty in turn-taking, especially in detecting whether the user was willing to start speaking and whether the user had finished speaking. This is related to the first issue; we need to use much richer information about multi-modality for better interaction.
- Emotion is an important issue in chat-oriented dialogue systems. Our system was not aware of user emotion, but we encountered cases in which users were not willing to continue with the current conversational topic. In such cases, it will be necessary to detect the emotion of users and change the current topic appropriately.
- In our argumentative dialogue system, it was rather difficult for humans to continue the discussion smoothly, even though we had large-scale argumentation structures. We believe this is mainly due to the difference in mental models between the system and humans. We need to find ways for humans and a system to have common conceptions and build common ground (Clark et al., 1991) so that discussion participants can build arguments on what has been discussed.

We are currently working with teams investigating paralinguistics and multi-modality to cope with the issues related to turn-taking and emotion. We are also considering

[2] https://www.youtube.com/watch?v=EpgBqjViyZE

ways for the system to disclose its personality, including its way of thinking, so that a common ground can be built and smooth discussion can be carried out.

3. Deployment of dialogue systems

Alongside our research, we have also been conducting field trials of dialogue systems, i.e., deployment of dialogue systems in the wild. We describe two case studies we conducted in Japan. The systems deployed are simple scenario-based systems so that it would be easy to customize them to make them fit actual environments and modify behaviors when necessary. In both cases, thousands of users used the deployed systems. We now describe the details of the field trials and their findings.

3.1. Case study 1

The first trial was conducted with NTT East Corporation and the Tokyo Chamber of Commerce. We placed Sota communication robots [3] on six different premises in Shinjuku, Tokyo, e.g., a fruit parlor, food company, book store, and department store. The robots were installed so that they could give guidance regarding the premises to their customers. The dialogue system is fully scenario-based. When the robot senses a customer with a human sensor, it addresses the customer and makes a greeting (opening phase). Then, the robot asks him/her if he/she had anything to ask about the premise. The system has a touch display to show the information asked by the customer (guidance phase). At the end of the interaction, the robot asked the customers for their level of satisfaction through a questionnaire and says good-bye to the user (closing phase). Figure 5 shows Sota on premises in the field trial.

For a period of four weeks, Sota attracted over 9,000 customers, out of which, about 4600 underwent the opening-phase of the dialogue (roughly one minute of interaction). About 4250 of these customers listened to the guidance from the robot, and about 1800 participated in the questionnaire at the closing phase. The averaged interaction time with the robot was just about one minute. Figure 6 shows the percentages of a three-scale evaluation (good, okay, bad evaluations) of the system through the questionnaire. When they reached the end of a dialogue, it seemed that many of the customers were satisfied with the system.

We asked the store owners/managers (N=14) about how they agreed with the following questionnaire items on a four-point Likert scale. The last question was answered with specific monetary values. Figures 7 and 8 show the results of the following questionnaire items:

Cost reduction The system contributed to the reduction in the cost (e.g., personnel expenses).

Sales increase The system contributed to an increase in sales.

PR effect The system had a positive PR effect.

Satisfaction It was a good idea to install the robot on my premise.

Future use I want to continue having the robot on my premise.

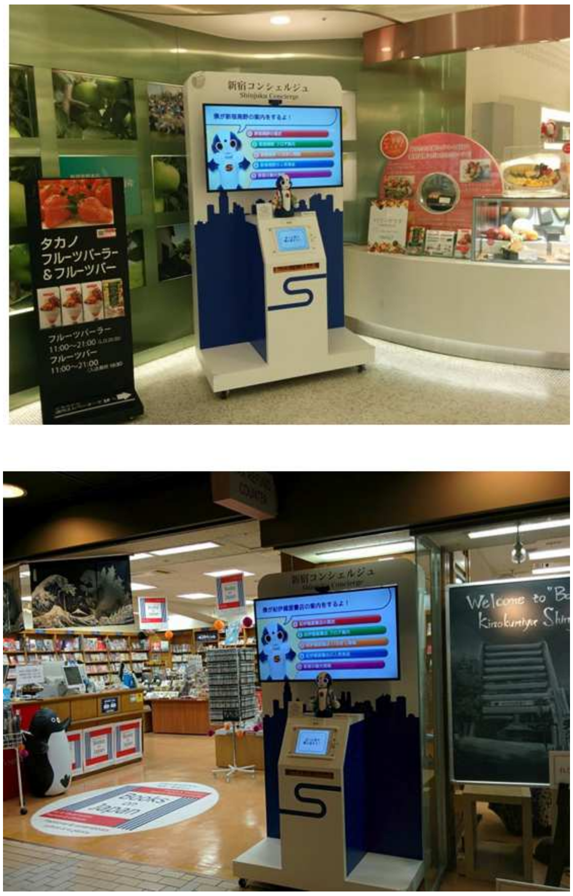

Figure 5: Sota on premises in Shinjuku, Japan

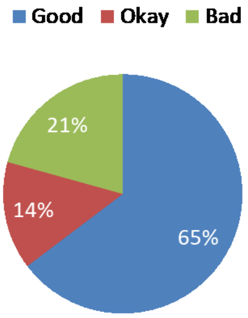

Figure 6: Questionnaire results from customers

Affordable cost How much can you afford to pay per month to have a robot on your premise? (for this item, N=13)

It can be seen that the store owners/managers were rather negative regarding the robot's effect on cost reduction and sales increase, although they felt it certainly had a positive PR effect. Overall, they were positive about having the robot on their premises and wanted to continue using it. One very interesting result was affordability. Most said they could only pay less than 30,000 yen (about 280 USD) per month, which is low compared to the cost of development, deployment, and maintenance.

[3] https://sota.vstone.co.jp/home/

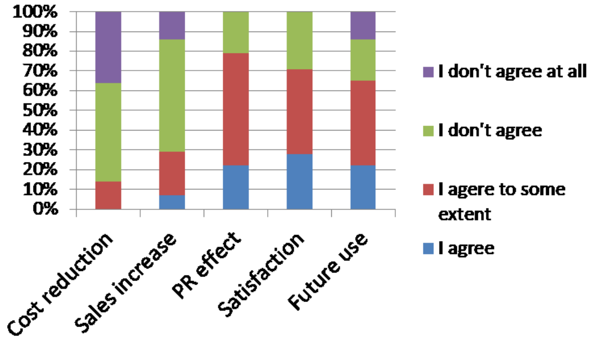

Figure 7: Questionnaire results from store owners/managers

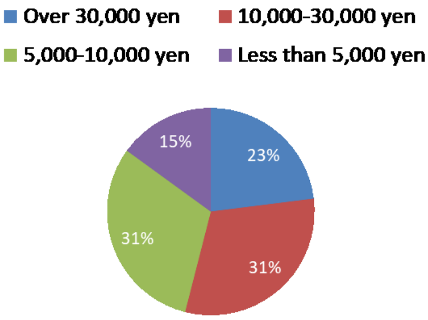

Figure 8: Questionnaire results regarding affordability (the price that owners/managers can afford for having the robot)

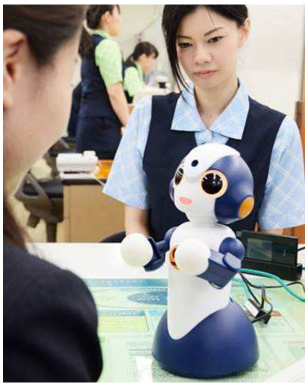

Figure 9: Sota at a regional bank

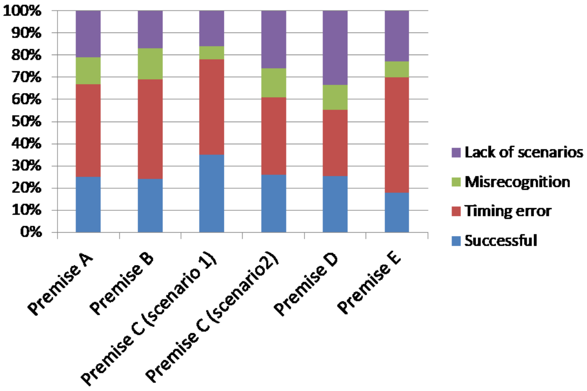

Figure 10: Interaction summary between Sota and customers

We encountered the following problems from this trial:

- The responsiveness of the robot should be improved. Speech-recognition accuracy is also a problem in actual noisy environments.
- The system has to cope with multiple languages of foreign customers. It is also necessary to cope with multiple customers at a time.
- The system needs to cope with nuances and emotional utterances.
- In addition to the information of the premises, the system was sometimes requested to provide information about neighboring areas and should cope with such requests.
- The system had limited information about the premises; it was necessary to show more detailed information when requested.

We learned many lessons from this trial. Although the system does not help from the sales point of view, the system was regarded to have some positive PR effect. Technically, the basic capability of the system needs to be improved, especially regarding responsiveness.

3.2. Case study 2

We conducted another trial involving several regional banks in Japan. This trial was conducted by NTT Data Corporation and several regional banks in Japan. Sota was installed on premises and interacted with customers to provide information. The robot could answer questions about housing loans, education loans, and other products by using a scenario. During a period of about four months, Sota interacted with over 8,000 customers, out of which several thousand engaged in verbal interaction with Sota. Figure 9 shows Sota interacting with a customer.

Figure 10 shows the summary of dialogues on five different premises, showing the percentage of successful (requested information was successfully provided to customer) and unsuccessful dialogues. For unsuccessful cases, the breakdown of the reasons (timing error, misrecognition, and lack of scenarios) are shown. It can be seen that the interactions were not very successful; about one fourth were successful. When we look at the breakdown of errors, we see that most of the errors were due to timing; the system could not respond to customers appropriately because it could not talk/listen to the customers at the right moment; when we listened to the recorded voices, we found that many were fragmented, with many initial parts stripped. This indicates that the customers started speaking, although the system was not ready for speech recognition. Compared to the timing issue, speech-recognition error was not a serious problem, although we should have prepared more scenarios to cope with more questions.

We encountered the following problems from this trial:

- The timing of the robot was the most serious issue. The customers were not aware of the robot's capability and interacted with the robot based on their sense of timing. The customers also had difficulty figuring out what they could do with the robot. It is necessary to explicitly state their functions, and if possible, the robot should act more proactively to provide information.
- The scenarios should be improved; it is necessary to add words/phrases and questions that were not included in the scenarios on a daily basis.

In this trial, we learned that timing was a primary issue with current dialogue systems when they are deployed on actual premises; this is in line with case study 1 in which we had an issue with responsiveness. In our research prototypes, we also had difficulty regarding timing when our chat-oriented/argumentative dialogue systems were built into androids. For deploying systems in the wild, timing has to be the primary concern.

4. Towards better timing

We have started working on multi-modal processing for better timing.

To make an utterance at an appropriate timing, it is necessary to estimate the end of an utterance of a user, queue of turn-taking from the user, and how long after the previous utterance to start speaking. The key is not only language information but also various nonverbal behaviors. For example, it is known that nonverbal behaviors, such as eye-gaze, head movement, breathing motion, and mouth movement, are useful in estimating the timing of turn-taking and appropriate utterances (Ishii et al., 2016a; Ishii et al., 2015; Ishii et al., 2016b; Ishii et al., 2016c; Ishii et al., 2017).

To estimate the appropriate timing more accurately, it is necessary to focus on more diverse nonverbal behaviors. In addition, there are many individual differences in nonverbal behavior depending on personality. There has not been sufficient research on the relationship between such nonverbal behavior and personal characteristics. To clarify the relationship between the proper timing of utterance and various and detailed nonverbal behaviors and to deal with personal characteristics and nonverbal behaviors, we are working on building a multi-modal corpus including various nonverbal behaviors and personal characteristics.

To construct a Japanese-conversation corpus including verbal and nonverbal behaviors in dialogue, we recorded 24 face-to-face two-person conversations (12 groups of two different people). The participants were Japanese males and females in their 20s to 50s who had never met before. They sat facing each other (Figure 11).

To acquire data of various dialogue scenes, three dialog scenes, i.e., discussion, chat, and story-telling, were recorded. In the story-telling scene, the participants had not seen the conversational content. Before the dialogue, they watched a famous popular cartoon animation called “Tom & Jerry” in which the characters do not speak. In each dialogue, one participant explained the content of the animation to the conversational partner. In each group, one session of discussion and chat and two sessions of story-telling were carried out.

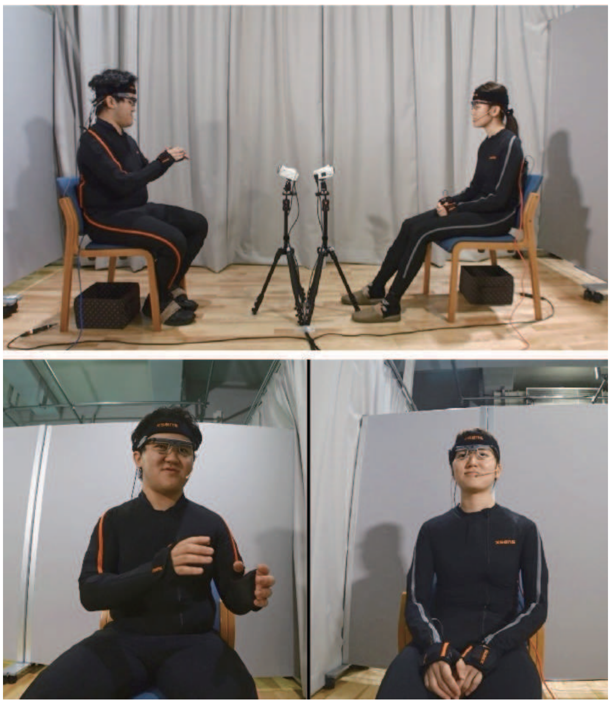

Figure 11: Two participants having dialogue

We recorded the participants' voices with a pin microphone attached to the chest and videoed the entire discussion. We also took bust (chest, shoulders, and head) shots of each participant (recorded at 30 Hz). In each dialogue, the data on the utterances and nodding behaviors of the person explaining the animation were collected in the first half of the ten-minute period (480 minutes in total) as follows.

- Utterances: We built an utterance unit using the inter-pausal unit (IPU) (Koiso et al., 1998). The utterance interval was manually extracted from the speech wave. A portion of an utterance followed by 200 ms of silence was used as the unit of one utterance.
- Gaze: The participants wore a glass-type eye tracker (Tobii Glass2). The gaze target of the participants and the pupil diameter were measured at 30 Hz.
- Body motion: The participants' body movements, such as hand gestures, upper body, and leg movements, were measured with a motion capture device (Xsens MVN) at 240 Hz.
- Personal trait: We obtained Big Five personality scores of the participants through subjective evaluation from the participants and a third party.

All verbal and nonverbal behavior data were integrated at 30 Hz for display using the ELAN viewer (Wittenburg et al., 2006). This viewer enabled us to annotate the multi-modal data frame-by-frame and observe the data intuitively.

In the future, we will clarify the relationship between the proper timing of an utterance and various and detailed nonverbal behaviors. We also want to deal with personal characteristics.

5. Summary and future work

We presented our research on chat-oriented and argumentative dialogue systems. We also described two case studies, one on various premises in Tokyo and the other in regional banks; we found that it is still a premature phase for systems to reduce cost or increase sales, but it seems that they have a positive PR effect. The current main problem of dialogue systems, in research and deployment alike, is timing. To this end, we started to work on multi-modal processing so that a system and users can interact more smoothly.

6. Acknowledgments

We would like to thank NTT East Corporation and Tokyo Chamber of Commerce for sharing the results of field trials. We also thank the members of NTT Data Corporation and the banks who participated in the trial for providing valuable data.

7. Bibliographical References

Clark, H. H., Brennan, S. E., et al. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149.

Gordon, T. F., Prakken, H., and Walton, D. (2007). The carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10):875–896.

Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y. (2014). Towards an open-domain conversational system fully based on natural language processing. In *Proc. COLING*, pages 928–939.

Higashinaka, R., Meguro, T., Sugiyama, H., Makino, T., and Matsuo, Y. (2015). On the difficulty of improving hand-crafted rules in chat-oriented dialogue systems. In *Proc. APSIPA*, pages 1014–1018.

Higashinaka, R., Funakoshi, K., Inaba, M., Tsunomori, Y., Takahashi, T., and Kaji, N. (2017a). Overview of dialogue breakdown detection challenge 3. In *Proc. Dialog System Technology Challenges Workshop (DSTC6)*.

Higashinaka, R., Sakai, K., Sugiyama, H., Narimatsu, H., Arimoto, T., Fukutomi, T., Matsui, K., Ijima, Y., Ito, H., Araki, S., Yoshikawa, Y., Ishiguro, H., and Matsuo1, Y. (2017b). Argumentative dialogue system based on argumentation structures. In *Proc. SemDial*, pages 154–155.

Imamura, K., Higashinaka, R., and Izumi, T. (2014). Predicate-argument structure analysis with zero-anaphora resolution for dialogue systems. In *Proc. COLING*, pages 806–815.

Ishii, R., Kumano, S., and Otsuka, K. (2015). Predicting next speaker based on head movement in multi-party meetings. In *Proc. ICASSP*, pages 2319–2323.

Ishii, R., Kumano, S., and Otsuka, K. (2016a). Analyzing mouth-opening transition pattern for predicting next speaker in multi-party meetings. In *Proc. ICMI*, pages 209–216.

Ishii, R., Otsuka, K., Kumano, S., and Yamamoto, J. (2016b). Predicting of who will be the next speaker and when using gaze behavior in multiparty meetings. *The ACM Transactions on Interactive Intelligent Systems*, 6(1):4.

Ishii, R., Otsuka, K., Kumano, S., and Yamamoto, J. (2016c). Using respiration to predict who will speak next and when in multiparty meetings. *The ACM Transactions on Interactive Intelligent Systems*, 6(2):20.

Ishii, R., Kumano, S., and Otsuka, K. (2017). Prediction of next-utterance timing using head movement in multiparty meetings. In *Proc. HAI*, pages 181–187.

Koiso, H., Horiuchi, Y., Tutiya, S., and Akira Ichikawa, a. Y. D. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs. *Language and Speech*, 41:295–321.

Koiso, H., Tsuchiya, T., Watanabe, R., Yokomori, D., Aizawa, M., and Den, Y. (2016). Survey of conversational behavior: Towards the design of a balanced corpus of everyday japanese conversation. In *Proc. LREC*.

Lippi, M. and Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, 16(2):10:1–10:25.

Meguro, T., Higashinaka, R., Minami, Y., and Dohsaka, K. (2010). Controlling listening-oriented dialogue using partially observable Markov decision processes. In *Proc. COLING*, pages 761–769.

Miltsakaki, E., Prasad, R., Joshi, A. K., and Webber, B. L. (2004). The Penn Discourse Treebank. In *Proc. LREC*.

Onishi, K. and Yoshimura, T. (2014). Casual conversation technology achieving natural dialog with computers. *NTT DOCOMO Technical Jouranl*, 15(4):16–21.

Otsuka, A., Hirano, T., Miyazaki, C., Higashinaka, R., Makino, T., and Matsuo, Y. (2017). Utterance selection using discourse relation filter for chat-oriented dialogue systems. In *Dialogues with Social Robots*, pages 355–365. Springer.

Sakai, K., Inago, A., Higashinaka, R., Yoshikawa, Y., Ishiguro, H., and Tomita, J. (2018). Creating large-scale argumentation structures for dialogue systems. In *proc. LREC*.

Shang, L., Sakai, T., Lu, Z., Li, H., Higashinaka, R., and Miyao, Y. (2016). Overview of the NTCIR-12 short text conversation task. In *Proc. NTCIR*.

Takeuchi, S., Cincarek, T., Kawanami, H., Saruwatari, H., and Shikano, K. (2007). Construction and optimization of a question and answer database for a real-environment speech-oriented guidance system. In *Proc. Oriental COCOSDA*, pages 149–154.

Vinyals, O. and Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Wallace, R. S. (2009). The anatomy of alice. In *Parsing the Turing Test*, pages 181–210. Springer.

Walton, D. (2013). *Methods of argumentation*. Cambridge University Press.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN a professional framework for multimodality research. In *Proc. LREC*.