Analysis of Listening-oriented Dialogue for Building Listening Agents

Toyomi Meguro, Ryuichiro Higashinaka, Kohji Dohsaka, Yasuhiro Minami, Hideki Isozaki

NTT Communication Science Laboratories, NTT Corporation

2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

{meguro,rh,dohsaka,minami,isozaki}@cslab.kecl.ntt.co.jp

Abstract

Our aim is to build listening agents that can attentively listen to the user and satisfy his/her desire to speak and have himself/herself heard. This paper investigates the characteristics of such listening-oriented dialogues so that such a listening process can be achieved by automated dialogue systems. We collected both listening-oriented dialogues and casual conversation, and analyzed them by comparing the frequency of dialogue acts, as well as the dialogue flows using Hidden Markov Models (HMMs). The analysis revealed that listening-oriented dialogues and casual conversation have characteristically different dialogue flows and that it is important for listening agents to self-disclose before asking questions and to utter more questions and acknowledgment than in casual conversation to be good listeners.

1 Introduction

Although task-oriented dialogue systems have been actively researched over the years (Walker et al., 2001), systems that perform more flexible (less task-oriented) dialogues such as chats are beginning to be actively investigated from their social and entertainment aspects (Bickmore and Cassell, 2001; Higuchi et al., 2008).

This paper deals with dialogues in which one conversational participant attentively listens to the other (hereafter, listening-oriented dialogue). Our aim is to build listening agents that can implement such a listening process so that a user can satisfy his/her desire to speak and have him/herself heard. Such agents would lead the user's state of mind for the better as in a therapy session, although we want our listening agents to help users mentally in everyday conversation. It should also be noted that the purpose of the listening-oriented dialogue is to simply listen to users, not to elicit information as in interviews.

L:	The topic is "travel", so did you travel during summer vacation?	(QUESTION)
S:	I like traveling.	(SELF-DISCLOSURE)
L:	Oh! I see!	(SYMPATHY)
	Why do you like to travel?	(QUESTION)
S:	This summer, I just went back to my hometown.	(SELF-DISCLOSURE)
	I was busy at work, but I'm planning to go to Kawaguchi Lake this weekend.	(SELF-DISCLOSURE)
	I like traveling because it is stimulating.	(SELF-DISCLOSURE)
L:	Going to unusual places changes one's perspective, doesn't it?	(SYMPATHY)
	You said you're going to go to Kawaguchi Lake this weekend. Is this travel?	(QUESTION)
	Will you go by car or train?	(QUESTION)

Figure 1: Excerpt of a typical listening-oriented dialogue. Dialogue acts corresponding to utterances are shown in parentheses (See Section 3.1 for their meanings). The dialogue was originally in Japanese and was translated by the authors.

There has been little research on listening agents. One exception is (Maatman et al., 2005), which showed that systems can make the user have the sense of being heard by using gestures, such as nodding and shaking of the head. Although our work is similar to theirs, the difference is that we focus more on verbal communication instead of non-verbal one.

For the purpose of gaining insight into how to build our listening agents, we collected listening-oriented dialogues as well as casual conversation, and compared them in order to reveal the characteristics of the listening-oriented dialogue. Figure 1 shows an example of a typical listening-oriented dialogue. In the figure, the conversational participants talk about travel with the listener (L), repeatedly asking the speaker (S) to make self-disclosure.

2 Approach

We analyze the characteristics of listening-oriented dialogues by comparing them with casual conversation. Here, casual conversation means a dialogue where conversational participants have no predefined roles (i.e., listeners and speakers). In this

study, we collect dialogues in texts because we want to avoid the particular problems of voice, such as filled pauses and interruptions, although we plan to deal with speech input in the future.

As a procedure, we first collect listening-oriented dialogues and casual conversation using human subjects. Then, we label the collected dialogues with dialogue act tags (see Section 3.1 for details of the tags) to facilitate the analysis of the data. In the analysis, we examine the frequency of the tags in each type of dialogue. We also look into the difference of dialogue flows by modeling each type of dialogue by Hidden Markov Models (HMMs) and comparing the obtained models. We employ HMMs because they are useful for modeling sequential data especially when the number of states is unknown. We check whether the HMMs for the listening-oriented dialogue and casual conversation can be successfully distinguished from each other to see if the listening process can be successfully modeled. We also analyze the transitions between states in the created HMMs to examine the dialogue flows. We note that HMMs have been used to model task-oriented dialogues (Shirai, 1996) and casual conversation (Isomura et al., 2006). In this study, we use HMMs to model and analyze listening-oriented dialogues.

3 Data collection

We recruited 16 participants. Eight participated as listeners and the other eight as speakers. The male-to-female ratio was even. The participants were 21 to 29 years old. Each participant engaged in four dialogues: two casual conversations followed by two listening-oriented dialogues with a fixed role of listener/speaker. In listening-oriented dialogue, the listeners were instructed to make it easy for the speakers to say what they wanted to say. When collecting the casual conversation, listeners were not aware that they would be listeners afterwards. Listeners had never met nor talked to the speakers prior to the data collection. The listeners and speakers talked over Microsoft Live Messenger[TM] in different rooms; therefore, they could not see each other.

In each conversation, participants chatted for 30 minutes about their favorite topic that they selected from the topic list we prepared. The topics were food, travel, movies, music, entertainers, sports, health, housework and childcare, personal computers and the Internet, animals, fashion and games. Table 1 shows the number of collected dialogues, utterances and words in each utterance of listeners and

		Listening	Casual
# dialogues		16	16
# utterances		850	720
# words per utt.	Listener	20.60	17.92
	Speaker	26.46	21.44

Table 1: Statistics of collected dialogues.

speakers. Generally, utterances in listening-oriented dialogue were longer than those in casual conversation, probably because the subjects explained themselves in detail to make themselves better understood.

At the end of each dialogue, the participants filled out questionnaires that asked for their satisfaction levels of dialogue, as well as how well they could talk about themselves to their conversational partners on the 10-point Likert scale. The analysis of the questionnaire results showed that, in listening-oriented dialogue, speakers were having a better sense of making themselves heard than in casual conversation (Welch's pairwise t-test; $p=0.016$) without any degradation in the satisfaction level of dialogue. This indicates that the subjects were successfully performing attentive listening and that it is meaningful to investigate the characteristics of the collected listening-oriented dialogues.

3.1 Dialogue act

We labeled the collected dialogues using the dialogue act tag set: (1) SELF-DISCLOSURE (disclosure of one's preferences and feelings), (2) INFORMATION (delivery of objective information), (3) ACKNOWLEDGMENT (encourages the conversational partner to speak), (4) QUESTION (utterances that expect answers), (5) SYMPATHY (sympathetic utterances and praises) and, (6) GREETING (social cues to begin/end a dialogue).

We selected these tags from the DAMSL tag set (Jurafsky et al., 1997) that deals with general conversation and also from those used to label therapy conversation (Ivey and Ivey, 2002). Since our work is still preliminary, we selected only a small number of labels that we thought were important for modeling utterances in our collected dialogues, although we plan to incorporate other tags in the future. We expected that self-disclosure would occur quite often in our data because the subjects were to talk about their favorite topics and the participants would be willing to communicate about their experiences and feelings. We also expected that the listeners would sympathize often to make others talk with ease. Note that sympathy has been found useful to increase closeness between conversational partic-

	Listener		Speaker	
	Casual	Listening	Casual	Listening
DISC	66.6%	44.5%	53.3%	57.3%
INFO	6.5%	1.4%	5.6%	5.2%
ACK	8.0%	12.3%	6.6%	6.9%
QUES	4.1%	25.8%	21.3%	14.0%
SYM	2.6%	3.7%	3.2%	3.3%
GR	10.9%	9.8%	7.2%	9.6%
OTHER	1.3%	2.5%	2.9%	3.7%

Table 2: Rates of dialogue act tags.

	DISC	INFO	ACK	QUES	SYM	GR
Increase	0	0	8	8	5	4
Decrease	8	8	0	0	3	4

Table 3: Number of listeners whose tags increased/decreased in listening-oriented dialogue.

ipants (Reis and Shaver, 1998).

A single annotator, who is not one of the authors, labeled each utterance using the seven tags (six dialogue act tags plus OTHER). As a result, 1,177 tags were labeled to the utterances in the listening-oriented dialogues and 1,312 tags to those in casual conversation. The numbers of tags and utterances do not match because, in text dialogue, an utterance can be long and may be annotated with several tags.

4 Analysis

4.1 Comparing the frequency of dialogue acts

We compared the frequency of the dialogue act tags in listening-oriented dialogues and casual conversation. Table 2 shows the rates of the tags in each type of dialogue. In the table, OTHER means the expressions that did not fall into any of our six dialogue acts, such as facial expressions and mistypes. Table 3 shows the number of listeners whose rates of tags increased or decreased from casual conversation to listening-oriented dialogue.

Compared to casual conversation, the rates of SELF-DISCLOSURE and INFORMATION decreased in the listening-oriented dialogue. On the other hand, the rates of ACKNOWLEDGMENT and QUES-TION increased. This means that the listeners tended to hold the transmission of information and focused on letting speakers self-disclose or deliver information. It can also be seen that the speakers decreased QUESTION to increase self-disclosure.

4.2 Modeling dialogue act sequences by HMM

We analyzed the flow of listening-oriented dialogue and casual conversation by modeling their dialogue act sequences using HMMs. We defined 14 observation symbols, corresponding to the seven tags for a listener and the same number of tags for a speaker.
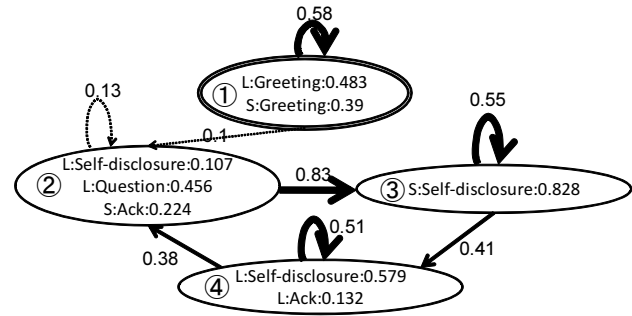

Figure 2: Ergodic HMM for listening-oriented dialogue. Circled numbers represent state IDs.

We trained the following two types of HMMs for each type of dialogue.

Ergodic HMM: Each state emits all 14 observation symbols. All states are connected to each other.

Speaker HMM: Half the states in this HMM only emit one speaker's dialogue acts and the other half emit other speaker's dialogue acts. All states are connected to each other.

The EM algorithm was used to train the HMMs. To find the best fitting HMM with minimal states, we trained 1,000 HMMs for each type of HMM by increasing the number of states from one to ten and training 100 HMMs for each number of states. This was necessary because the HMMs severely depend on the initial probabilities. From the 1,000 HMMs, we chose the most fitting model using the MDL (Minimum Description Length) criterion.

4.2.1 Distinguishing Dialogue Types

We performed an experiment to examine whether the trained HMMs can distinguish listening-oriented dialogues and casual conversation. For this experiment, we used eight listening-oriented dialogues and eight casual conversations to train HMMs and made them classify the remaining 16 dialogues. We found that Ergodic HMM can distinguish the dialogues with an accuracy of 87.5%, and the Speaker HMM achieved 100% accuracy. This indicates that we can successfully train HMMs for each type of dialogue and that investigating the trained HMMs would show the characteristics of each type of dialogue. In the following sections, we analyze the HMMs trained using all 16 dialogues of each type.

4.2.2 Analysis of Ergodic HMM

Figure 2 shows the Ergodic HMM for listening-oriented dialogue. It can be seen that the major flow
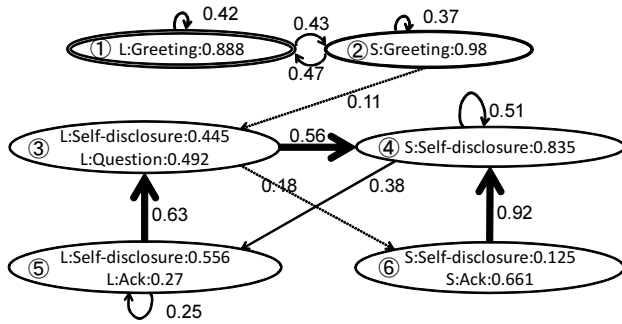
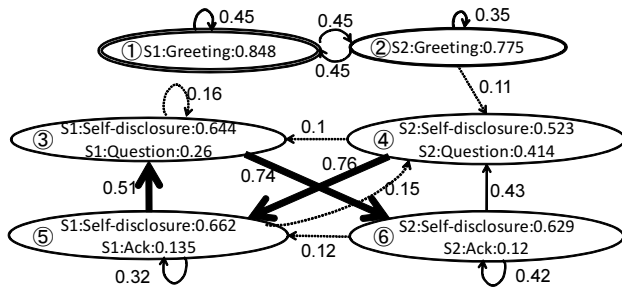Figure 3: Speaker HMM for listening-oriented dialogue.

Figure 4: Speaker HMM for casual conversation.

of dialogue acts are: ② L's question → ③ S's self-disclosure → ④ L's self-disclosure → ② L's question. This flow indicates that listeners tend to self-disclose before the next question, showing the cycle of reciprocal self-disclosure. This indicates that listening agents would need to have the capability of self-disclosure in order to become human-like listeners.

4.2.3 Analysis of Speaker HMM

Figures 3 and 4 show the Speaker HMMs for listening-oriented dialogue and casual conversation, respectively. Here, L and S correspond to S1 and S2. It can be clearly seen that the two HMMs have very similar structures. From the probabilities, states with the same IDs seem to correspond to each other. When we compare state IDs 3 and 5, it can be seen that, when speakers take the role of listeners, they reduce self-disclosure while increasing questions and acknowledgment. Questions seem to have more importance in listening-oriented dialogue than in casual conversation, indicating that listening agents need to have a good capability of generating questions. The agents would also need to explicitly increase acknowledgment in their utterances. Note that, compared to spoken dialogue, acknowledgment has to be performed consciously in text-based dia-

logue. When we compare state ID 4, we see that the speaker starts questioning in casual conversation, whereas the speaker only self-discloses in listening-oriented dialogue. This shows that, in our data, the speakers are successfully concentrating on making self-disclosure in listening-oriented dialogue.

5 Conclusion and Future work

We collected listening-oriented dialogue and casual conversation, and compared them to find the characteristics of listening-oriented dialogues that are useful for building automated listening agents. Our analysis found that it is important for listening agents to self-disclose before asking questions and that it is necessary to utter more questions and acknowledgment than in casual conversation to be good listeners. As future work, we plan to use a more elaborate tag set to further analyze the dialogue flows. We also plan to extend the HMMs to Partially Observable Markov Decision Processes (POMDPs) (Williams and Young, 2007) to achieve dialogue management of listening agents from data.

References

Timothy Bickmore and Justine Cassell. 2001. Relational agents: A model and implementation of building user trust. In *Proc. ACM CHI*, pages 396–403.

Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. 2008. A casual conversation system using modality and word associations retrieved from the web". In *EMNLP*, pages 382–390.

Naoki Isomura, Fujio Toriumi, and Kenichiro Ishii. 2006. Evaluation method of non-task-oriented dialogue system by HMM. In *Proc. the 4th Symposium on Intelligent Media Integration for Social Information Infrastructure*, pages 149–152.

Allen E. Ivey and Mary Bradford Ivey. 2002. *Intentional Interviewing and Counseling: Facilitating Client Development in a Multicultural Society*. Brooks/Cole Publishing Company.

Dan Jurafsky, Liz Shriberg, and Debra Biasca, 1997. *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual*.

Martijn Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. *Lecture Notes in Computer Science*, 3661:25–36.

Harry T. Reis and Phillip Shaver. 1998. Intimacy as an interpersonal process. In S. Duck, editor, *Handbook of personal relationships*, pages 367–398. John Wiley & Sons Ltd.

Katsuhiko Shirai. 1996. Modeling of spoken dialogue with and without visual information. In *Proc. ICSLP*, volume 1, pages 188–191.

Marilyn A. Walker, Rebecca Passonneau, and Julie E. Boland. 2001. Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems. In *Proc. ACL*, pages 515–522.

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