Can We Predict Who in the Audience will Ask What Kind of Questions with their Feedback Behaviors in Poster Conversation?

Tatsuya Kawahara, Takuma Iwatate, Takanori Tsuchiya, Katsuya Takanashi

School of Informatics, Kyoto University,
Sakyo-ku, Kyoto 606-8501, Japan

Abstract

We investigate feedback behaviors in conversations in poster sessions, specifically whether it is possible to predict who in the audience will ask questions, and also what kind of questions. We focus on verbal backchannels and non-verbal noddings by the audience as well as joint eye-gaze events by the presenter and the audience. We first show how these patterns are correlated with turn-taking by the audience. Then, questions made by the audience are classified into two kinds: confirming questions and substantive questions. It is suggested that only verbal backchannels are useful for distinguishing them.

1. Introduction

Feedback behaviors are important cues in analyzing presentation-style conversations. We can guess whether the audience is attracted to the presentation by observing their feedback behaviors. This characteristic is more prominent when the audience is smaller; the audience can make not only non-verbal feedbacks such as nodding, but also verbal backchannels. We have been collecting and analyzing poster conversations, in which a researcher makes an academic presentation to a couple of persons using a poster. In our previous work [1], we demonstrated that non-lexical kinds of verbal backchannels, referred to as reactive tokens, are a good indicator of the audience's interest level.

Poster sessions have become a norm in many academic conventions because of the interactive characteristics. The audience can ask questions even during the presentation. We can also guess whether the presentation is understood or liked by the audience by observing the quantity and quality of their questions. It is also known that the turn-taking behavior is related with the backchannel and gaze patterns [2, 3, 4]. The goal of this work is to investigate the relationship of these feedback behaviors with turn-taking by the audience and also the kind of questions they ask. We classify the audience's questions into two kinds: confirming questions and substantive questions. We expect that these analyses reveal how the audience appreciates the presentation and the quality of the poster conversation.

2. Multi-modal Corpus of Poster Conversations

We have recorded a number of poster conversations designed for multi-modal data collection [5]. In this study, we use four poster sessions, in which the presenters and audiences are different from each other. In each session, one presenter (labeled as "A") prepared a poster on his/her own academic research, and there was an audience of two persons (labeled as "B" and "C"), standing in front of the poster and listening to the presentation. They were not familiar with the presenter and had not heard the presentation before. The duration of each session was 20-30 minutes.

All speech data were segmented into IPUs (Inter-Pausal Unit) with time and speaker labels, and transcribed according to the guideline of the Corpus of Spontaneous Japanese (CSJ) [6]. We also manually annotated fillers and verbal backchannels.

The recording environment was equipped with multi-modal sensing devices such as cameras and a motion capturing system while every participant wore an eye-tracking recorder and an accelerometer attached with a cap. Noddings are detected with the accelerometer. Eye-gaze information is derived from the eye-tracking recorder and the motion capturing system by matching the gaze vector against the position of the other participants and the poster.

An outlook of session recording is given in Figure 1.

3. Relationship between Feedback Behaviors and Turn-Taking

First, we investigate statistics of eye-gaze and backchannel events and their relationship with turn-taking by the audience.

3.1. Duration of Eye-gaze

We identify the object of the eye-gaze of all participants at the end of the presenter's utterances. The target object can be either the poster or other participants. Then, we measure the duration of the eye-gaze within the segment of 2.5 seconds before the end of the presenter's ut-

Figure 1: Outlook of poster session recording

Table 2: Definition of joint eye-gaze events by presenter and audience

who	presenter		
	gazes at	audience (**I**)	poster (**P**)
audience	presenter (**i**)	**Ii**	**Pi**
	poster (**p**)	**Ip**	**Pp**

Table 3: Statistics of joint eye-gaze events by presenter and audience in relation with turn-taking

	#turn held by presenter A	#turn taken by audience (self)	(other)	total
Ii	125	17	3	145
Ip	320	**71**	26	417
Pi	190	11	9	210
Pp	2974	147	145	3266

Table 1: Duration (sec.) of eye-gaze and its relationship with turn-taking

	turn held by presenter A	turn taken by audience B	C
A gazed at B	0.220	**0.589**	0.299
A gazed at C	0.387	0.391	**0.791**
B gazed at A	0.161	0.205	0.078
C gazed at A	0.308	0.215	0.355

terances because the majority of the IPUs are less than 2.5 seconds. It is listed in Table 1 in relation with the turn-taking events. We can see the presenter gazed at the person right before yielding the turn to him/her significantly longer than other cases. However, there is no significant difference in the duration of the eye-gaze by the audience according to the turn-taking events.

3.2. Joint Eye-gaze Events

Next, we define joint eye-gaze events by the presenter and the audience as shown in Table 2. In this table, we use notation of "audience", but actually these events are defined for each person in the audience. Thus, "Ii" means the mutual gaze by the presenter and a particular person in the audience, and "Pp" means the joint attention to the poster object.

Statistics of these events at the end of the presenter's utterances are summarized in Table 3. Here, the counts of the events are summed over the two persons in the audience. They are classified according to the turn-taking events, and turn-taking by the audience is classified into two cases: the person involved in the eye-gaze event actually took the turn (self), and the other person took the turn (other). The mutual gaze ("Ii") is expected to be related with turn-taking, but its frequency is not so high. The frequency of "Pi" is not high, either. The most potentially useful event is "Ip", in which the presenter gazes

at the person in the audience before giving the turn. This is consistent with the observation in the previous subsection.

3.3. Backchannels

Verbal backchannels, typically "*hai*" in Japanese and "yeah" or "okay" in English, indicate the listener is understanding what is being said. They also suggest the listener's interest level [7, 1] and activate interaction. Nodding is regarded as a non-verbal backchannel, and it is more frequently observed in poster conversations than in simple spoken dialogues.

The occurrence frequencies of these events are counted within the segment of 2.5 seconds before the end of the presenter's utterances. They are shown in Figure 2 according to the joint eye-gaze events. It is observed that the person in the audience who takes the turn (=turn-taker) made more backchannels both in verbal and non-verbal manners, and the tendency is more apparent in the particular eye-gaze events of "Ii" and "Ip" which are closely related with the turn-taking events.

3.4. Discussions

It is shown that the most relevant features among the eye-gaze information is the presenter's gazing at the person to whom the turn is to be yielded. This is presumably affected by the characteristics of the poster conversation in which the presenter takes a major role in the conversation. The backchannel information by the audience may also be effective in turn-taking. The feedback not only indicates the audience's reaction, but also will attract the presenter's attention, triggering his/her gazing and then turn-yielding.
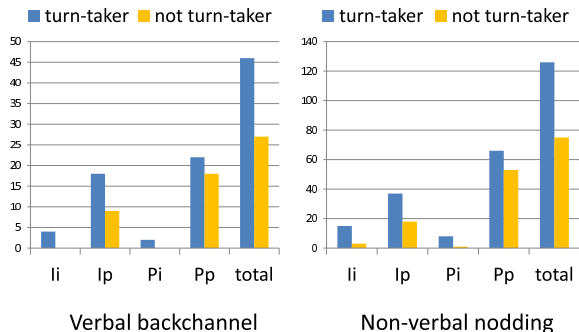
Figure 2: Statistics of backchannels and their relationship with turn-taking

4. Relationship between Feedback Behaviors and Kind of Questions

Next, we investigate the relationship between feedback behaviors of the audience and the kind of questions they ask after they take a turn. In this work, questions are classified into confirming questions and substantive questions. The confirming questions are asked to make sure of the understanding of the current explanation, thus they can be answered simply by "Yes" or "No".[1] The substantive questions, on the other hand, are asking about what was not explained by the presenter, thus they cannot be answered by "Yes" or "No" only; an additional explanation is needed.

This annotation together with the preceding explanation segment is not so straightforward when the conversation got into the QA phase after the presenter went through an entire poster presentation. Thus, we exclude the QA phase and focus on the questions made during the explanation phase. In this section, we analyze the behaviors during the explanation segment that precedes the question by merging all consecutive IPUs of the presenter. This is a reasonable assumption once turn-taking is predicted in the previous section. These are major differences from the analysis of the previous section.

4.1. Backchannels

The occurrence frequencies of verbal backchannels and non-verbal noddings, normalized by the duration of the explanation segment (seconds), are listed according to the question type in Tables 4 and 5. In these tables, statistics of the person who actually asked questions are compared with those of the person who did not. We can observe the turn-taker made significantly more verbal backchannels when asking substantive questions. On the other hand, there is no significant difference in the frequency of non-

Table 4: Frequencies (per sec.) of verbal backchannels and their relationship with question type

	confirming	substantive
turn-taker	0.034	**0.063**
non-turn-taker	0.041	0.038

Table 5: Frequencies (per sec.) of non-verbal noddings and their relationship with question type

	confirming	substantive
turn-taker	0.111	0.127
non-turn-taker	0.109	0.132

Table 6: Duration (ratio) of joint eye-gaze events and their relationship with question type

	confirming	substantive
Ii	0.053	0.015
Ip	**0.116**	0.081
Pi	0.060	0.035
Pp	0.657	**0.818**

verbal noddings among the audience and among the question types.

4.2. Eye-gaze Events

We also investigate the relationship between eye-gaze events and the question type. Among several parameterizations introduced in the previous section, we observe a significant tendency in the duration of the joint eye-gaze events, which is normalized by the duration of the presenter's explanation segment. It is summarized in Table 6. We can see the increase of "Ip" (and decrease of "Pp" accordingly) in confirming questions. By combining with the analysis in the previous section, we can reason the majority of turn-taking signaled by the presenter's gazing is attributed to confirmation.

4.3. Discussions

When the audience asks substantive questions, the presentation and understanding should be into a deep level. Thus, both the presenter and the audience are focused on the poster. The verbal backchannels might signal the person's confidence and interest level, but we need to investigate the syllabic and prosodic patterns of these reactive tokens as in the previous work [8].

5. Conclusions

We have investigated the relationship between feedback behaviors of the audience and the succeeding events of making questions in poster conversations. It is confirmed that the gaze information plays an important role in turn-

---

[1]This does not mean the presenter actually answered simply by "Yes" or "No".

taking, but it is more relevant to confirming questions rather than substantive questions. Both verbal backchannels and non-verbal noddings are also correlated with the turn-taking events, but verbal backchannels are more relevant to substantive questions. We presume that these different feedback behaviors are related with the understanding or interest level of the audience.

Based on the findings, we plan to design a smart posterboard which can control cameras and a microphone array to record the sessions and annotate the audience's reaction, which is critically important in poster conversations. These findings will also be useful for an intelligent conversational agent that makes an autonomous presentation.

6. References

[1] T.Kawahara, K.Sumi, Z.Q.Chang, and K.Takanashi. Detection of hot spots in poster conversations based on reactive tokens of audience. In *Proc. INTERSPEECH*, pages 3042–3045, 2010.

[2] Nigel G. Ward and Yaffa Al Bayyari. A Case Study in the Identification of Prosodic Cues to Turn-Taking: Back-Channeling in Arabic. In *Proc. INTERSPEECH*, pages 2018–2021, 2006.

[3] Bo Xiao, Viktor Rozgic, Athanasios Katsamanis, Brian R. Baucom, Panayiotis G. Georgiou, and Shrikanth Narayanan. Acoustic and Visual Cues of Turn-Taking Dynamics in Dyadic Interactions. In *Proc. INTERSPEECH*, pages 2441–2444, 2011.

[4] K.Jokinen, K.Harada, M.Nishida, and S.Yamamoto. Turn-alignment using eye-gaze and speech in conversational interaction. In *Proc. INTERSPEECH*, pages 2018–2021, 2011.

[5] T.Kawahara, H.Setoguchi, K.Takanashi, K.Ishizuka, and S.Araki. Multi-modal recording, analysis and indexing of poster sessions. In *Proc. INTERSPEECH*, pages 1622–1625, 2008.

[6] K.Maekawa. Corpus of Spontaneous Japanese: Its design and evaluation. In *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pages 7–12, 2003.

[7] T.Kawahara, M.Toyokura, T.Misu, and C.Hori. Detection of feeling through back-channels in spoken dialogue. In *Proc. INTERSPEECH*, page 1696, 2008.

[8] T.Kawahara, Z.Q.Chang, and K.Takanashi. Analysis on prosodic features of Japanese reactive tokens in poster conversations. In *Proc. Int'l Conf. Speech Prosody*, 2010.