Predicting Turn Taking from Gaze Transition Patterns Considering Participation Status in Multi-Party Conversation

Takashi Makino, Yoshinari Takegawa, and Keiji Hirata

***Abstract*—In multi-party conversation, listener's role is not uniform. Listeners are divided into those who are addressed by the speaker (addressee) and those who are not addressed by the speaker (side-participant). In this paper, we propose a mathematical model to predict whether turn-taking or turn-keeping occurs. For a feature quantity of a model, we focus on gaze transition patterns near the end of utterance for each role of a participant. We analyze the difference in frequencies of gaze transition patterns between addressee and side-participant. Based on the result of analysis, we construct a probabilistic model that considers listener's roles. As a result, the proposed model outperforms the conventional model that does not consider listener's roles.**

***Index Terms*—turn-taking, participation status, gaze, multi-party conversation.**

I. Introduction

PARTICIPANTS in multi-party conversation take speaking turns smoothly although those who speak and when ones speak are not determined in advance. Smooth communication involving turn-taking is investigated for conversational systems. There are many studies conducting analysis on turn-taking [8], [2]. In a two-person conversation, it is relatively easy to predict a turn-taking because a listener is necessarily always one person. On the other hand, in multi-party conversation, turn-taking is more complex than two-person conversation because there are possibly multiple listeners and multiple next speaker candidates. The participants must predict the timing when the speaker's utterance ends and to consider the start of timing for speaking when you become the next speaker in multi-party conversation. To build a conversational system which is able to speak at natural timing and produces appropriate gaze behaviors, the system has to be able to predict the timings like participants. Thus, we think that it is promising to analyze turn-taking and construct a model which can predict turn-taking.

In the field of social psychology, it has been shown that not only verbal information but also non-verbal information such as gaze information is regarded important for turn-taking prediction. Kendon [2] points out that there are tendencies between gaze behaviors and turn-taking. For example, when the turn-taking occurs, a next speaker makes mutual gaze with the speaker, and the listener who is not the next speaker gazes at the next speaker before the utterance of the next speaker is started. Moreover, based on the above knowledge, in multi-party conversation, the models using gaze information of a speaker and listeners to predict turn-taking and a next speaker have been studied [12], [17].

We argue that the listeners should be distinguished by roles although the previous models regard listeners identical. Goffman [4] splits conversation participants into ratified participants and overhearers according to roles. In addition, the former is also divided into speaker, addressee, and side-participant. In short, listener has two roles, addressee and side-participant. Lerner [6] points out that gaze behaviors are varied between addressee and side-participant when a speaker selects a next speaker.

We aim to construct a mathematical model to predict whether turn-taking or turn-keeping occurs with higher performance by considering listeners' roles. First, we analyze the relationship between gaze behaviors and listener's roles. Next, we analyze the relationship between gaze behaviors and whether turn-taking or turn-keeping occurs for each listeners' role. Based on the experimental results, we construct a prediction model considering listener's role. The result of the model evaluation shows that the proposed one outperforms the conventional model that does not take into account listeners' roles.

II. Related Work

A. Turn-taking in Social Psychology Researches

Most of the research related to unraveling the mechanism of the turn-taking in conversation has been executed in the field of social psychology. Sacks et al. [8] propose the turn-taking model. The model defines that there are transition relevance points (TRPs) near the end of the utterance. Especially, gaze behaviors are important in turn-taking. Kendon [2] claims that a speaker gazes at a listener as a turn-yielding cue to give a turn to the listener at the end of an utterance. In addition, he also status that a listener gazes at a speaker and looks away from the speaker as speaker-state signals when the listener accepts turn-taking. Jokinen [9] also insists on similar trends in multi-party conversation. Thus, it is important for turn-taking to focus on gaze behaviors.

In multi-party conversation, more complicated interaction occurs than two-person conversation because there are multiple listeners. To understand complicated interactions, it is important to focus on behavior of each role of participants. Goffman propose distinction of participation status in conversation [4] . Conversation participants are divided ratified participants and overhearers. Ratified participants are the participants who can become a speaker at any time. In addition, ratified participants are divided into speakers, addressees and

Manuscript received July 02, 2018; revised August 2, 2018.

T. Makino is with Future University Hakodate Graduate School, Japan, Hakodate, e-mail: g2117044@fun.ac.jp.

Y. Takegawa and K. Hirata are with Future University Hakodate, email: { yoshi, hirata }@fun.ac.jp

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

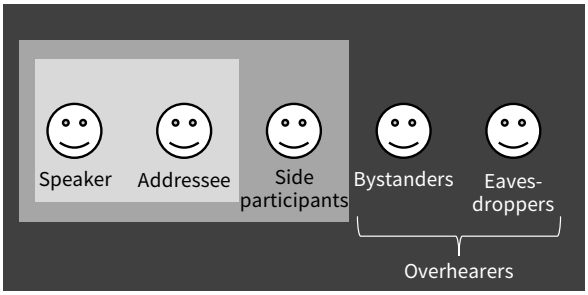

Fig. 1. Participation status

side-participants. Addressees are the participants are targeted by speaker. Side-participants are the others. Side-participants are allowed to join the conversation at any time although they are not expected to speak or response. Lerner [6] points out that there are conditions not only for speakers but also for addressees and side-participants in order to succeed in selecting the next speaker by gazing at a participant. (I) The addressee intended by the speaker needs to look at the speaker's gaze. (II) The side-participants also have to know that the other participant is the addressee by looking at the speaker's gaze. Thus, in multi-party conversation, the appropriate behaviors according to the roles such as not only speaker but also addressee and side-participant are observed when turn-taking occurs.

These studies support using gaze behavior of each role, speaker, addressee, and side-participant to predict turn-taking.

B. Prediction of Turn-taking in Engineering Researches

Several studies have carried out prediction of turn-taking by using relationship between turn-taking and non-verbal information. Some researchers have used speech processing to estimate whether turn-taking or turn-keeping occurs at the end of an utterance. For instance, Ferrer et al [10] use prosodic information, Schlangen et al [3] use vocabularies and prosodic information, Levow et al [5] use tonal language in order to estimate turn-taking. On the other hand, others have used gaze information to estimate turn-taking. Dielmann [1] uses physical motion, Jokinen [9] uses gaze information, and Ishii et al[13] use respiration information to detect turn-taking. Iwan de kok et al [11] propose the detection model of turn-taking using dialog acts, prosody, head-gesture and speaker gaze information. These studies have shown that gaze behavior is more important than speech information to estimate turn-taking.

In addition, there are several studies that conduct prediction models using more detailed gaze information. Kawahara et al [17] propose a turn-taking detection model using prosody and participants' gaze information, such as the person gazed upon and presence or absence of mutual gaze in a three-person poster conversation. Ishii et al [12] show that using a detailed transition pattern such as how the gaze target has changed and mutual gazes can detect turn-taking more accurately than using a single line of gaze. To treat gaze information for predicting turn-taking, detailed gaze information such as a mutual gaze and a transition pattern is more effective than a single line of gaze.

III. Corpus of multi-party conversation

We use twelve conversations from Chiba University three people conversation data [18] for analysis. This corpus is free conversation data of three people of similar sex with affinity. A topic in the conversations is selected by a dice at random. There are some topics such as a story of love and a smelly story. At that time, they are taught that they do not have to worry about the topic and that the topic may change on the way. The age of the participants is 18 to 33 years old, and they are all in Japanese. The duration of each session is about ten minutes.

A. Utterance Unit

All utterances are divided into Long Utterance Units (LUUs) [19]. LUU is designed as an analysis unit that contributes to researches such as analyzing mutual actions among conversation participants. Den et al [19] mention that the timing of turn-taking is near the end of LUU.

B. Gaze Object

Gaze objects are annotated for each participant based on the video data. The start point of the gaze object is the point at which the gaze starts to move to a certain participant from previous fixed point. The end of gaze object is the point at which the gaze starts to move to the other participants or other than participants. The starting points and the end points are checked by frame advance of the video data and annotated with an accuracy of one-thirtieth second.

C. Method of A for Listener's Role

We define listeners' roles based on participation status. Several studies have been pointed out that the gaze and body orientation greatly influence the determination of participation status. Duncan et at and Kendon show that turn-taking is adjusted by gaze and gesture [14], [15], [2]. Thus, we classified the roles of the two listeners as follows in order to analyze the difference between gaze behaviors and each listeners' role.

- Addressee: A person seen most frequently by speaker near the end of the current LUU.
- Side-participant: A person of a listener other than addressee.

In the case where the speaker gazes only other than participant at the end of the current LUU, the roles of both listeners are regarded as side-participants.

IV. Prediction model of Turn-taking

A. Gaze Transition Pattern

To analyze gaze behavior of each listeners' role, we use gaze transition patterns. Ishii et al [12] point out that it is important to focus on not only the gaze of the end of an utterance but also the gaze transition or the mutual gaze. Therefore, we decided to focus on the gaze transition patterns near the end of utterance. The transition pattern is explained below based on the definition of Ishii et al [12]. To generate a gaze transition pattern, we focus on the interval of 1200 ms from 1000ms before the end of the utterance to 200 ms after the end of it. We express temporal transitions of participants'

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

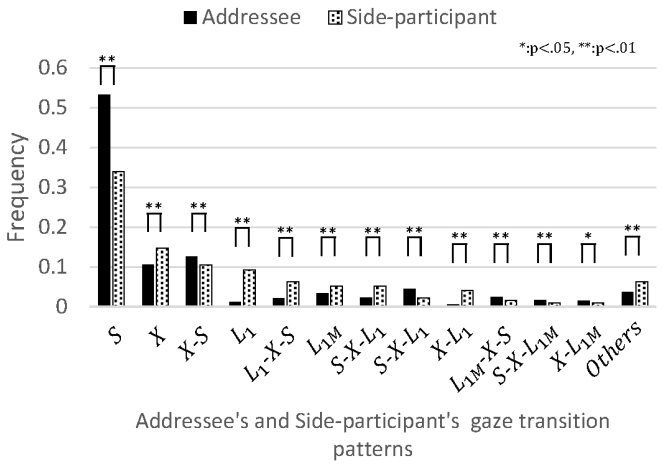

Fig. 2. Relationship between gaze transition patterns of addressees and side-participants

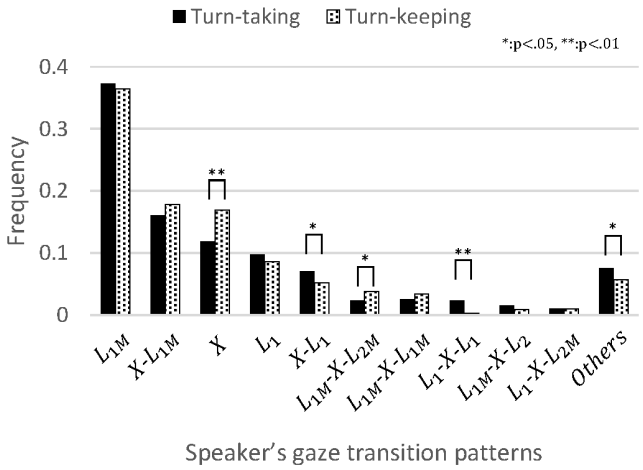

Fig. 3. Relationship between speaker's gaze transition patterns and turn-taking

gaze behaviors as n-gram (gazed object), which is define as a sequence of gaze direction shifts. The gazed objects were first classified as speaker, listener, others and labeled. We use the following five gaze labels for classification:

- S: A listener looks at a speaker without a mutual gaze.
- S_M: A listener looks at a speaker with a mutual gaze.
- L_1-L_2: A speaker or a listener looks at the other listeners without a mutual gaze. L_1 and L_2 show different listeners.
- L_{1M}-L_{2M}: A speaker or a listener looks at the other listeners with a mutual gaze. L_{1M} and L_{2M} show different listeners.
- X: A speaker or a listener looks at places other than participants, such as a floor or a ceiling.

B. Analysis of Gaze Transition Pattern and Participation Status

First, we analyze how quantitatively the change in the gaze transition pattern of participant differs by participation status. We introduce the results of analyzing gaze transition pattern of addressee and side-participant.

Fig. 2 shows the frequencies of gaze transition patterns for each addressees and side-participants using 7398 data. As a result , there are forty nine gaze transition patterns. The "Others" class includes forty one patterns, each of which are occurred in less than 1% of the data because the number of data is small. The result of a chi-squared test shows that the frequencies of gaze transition patterns differ significantly between addressee and side-participant ($\chi = 667.09$, $df = 12$, $p < .01$). Next, to verify which gaze transition patterns differ between addressee and side-participant, we conduct a residual analysis [16]. The result is shown in 2, from which we understand the following.

- Addressees' gaze transition patterns have significantly high frequencies at the time of S, X-S, S-X-S, L_{1M}-X-S, S-X-L_{1M}, and X-L_{1M}. That is, when a listener keeps looking at a speaker or the other listener, begins to gaze at a speaker, or begins a mutual gaze with the other listener, the frequency that a listener is an addressee is high.
- Side-participants' gaze transition patterns have significantly high frequencies at the time of X, L_1, L_{1M}, S-X-L_1, X-L_1, and $Others$. That is, when a listener does not look at the other participants at all, keeps the other listener, looks at both a speaker and the other listener, or continues a mutual gaze with the other listener, the frequency that a listener is a side-participant is high.

Therefore, these results show that gaze transition patterns are different according to listeners' roles.

C. Analysis of Gaze Transition Pattern and Turn-taking

To construct a prediction model of turn-taking, we analyze how much changes in the gaze transition pattern would differ quantitatively by a turn-taking and turn-keeping for each participation status.

Fig. 3 shows the frequencies of speakers' gaze transition pattens under turn-taking and turn-keeping conditions using 3699 data. As a result, there are fifty-one gaze transition patterns. The "Others" class includes thirty-three patterns, each of which are occurred in less than 1% of the data because the number of data is small. The result of a chi-squared test shows that the frequencies of gaze transition patterns differ significantly between turn-taking and turn-keeping ($\chi = 63.31$, $df = 11$, $p < .01$). Next, to verify which gaze transition patterns differ between conditions, we conduct a residual analysis [16]. The result is shown in 3, from which we understand the following.

- Speakers' gaze transition patterns have significantly high frequencies at the time of X-L_1, L_1-X-L_2, and $Others$. That is, when a speaker begins to look at a listener and looks at both listeners, the frequency of turn-taking is high.
- Speakers' gaze transition patterns have significantly high frequencies at the time of X and L_{1M}-X-L_{2M}. That is, when a speaker does not look listeners at all or looks at both listeners with mutual gazes, the frequency of turn-keeping is high.

Fig. 4 shows the frequencies of addressees' gaze transition patterns under turn-taking and turn-keeping conditions using

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

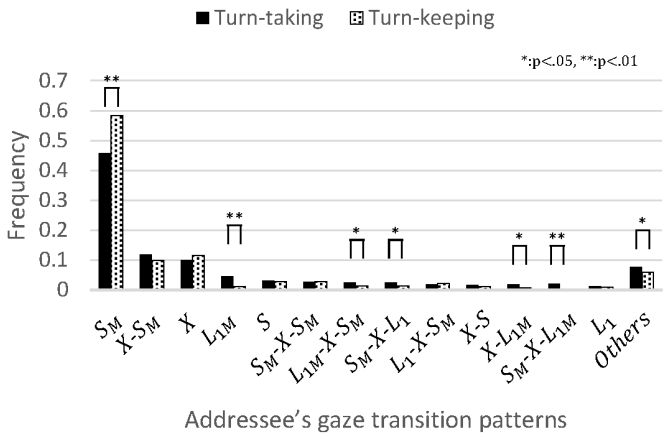

Fig. 4. Relationship between addressee's gaze transition patterns and turn-taking

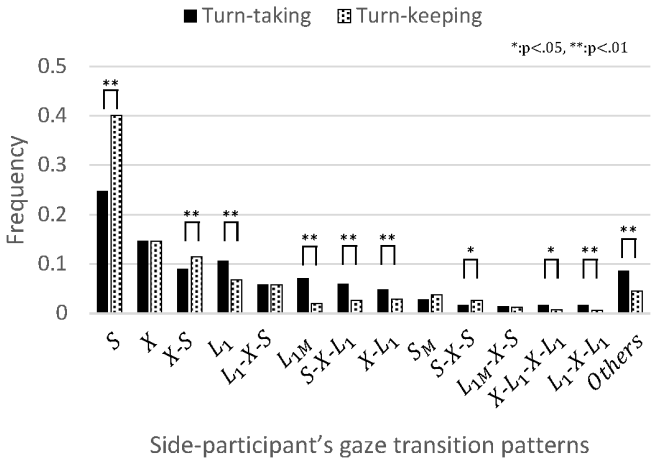

Fig. 5. Relationship between side-participant's gaze transition patterns and turn-taking

3189 data. As a result, there are sixty-eight gaze transition patterns. The "Others" class includes fifty-five patterns, each of which are occurred in less than 1% of the data because the number of data is small. The result of a chi-squared test shows that the frequencies of gaze patterns differ significantly between turn-taking and turn-keeping ($\chi = 90.15$, $df = 13$, $p < .01$). Next, to verify which gaze transition patterns differ between conditions, we conduct a residual analysis [16]. The result is shown in 4, from which we understand the following.

- Addressees' gaze transition patterns have significantly high frequencies at the time of L_{1M}, L_{1M}-X-S_M, S_M-X-L_1, X-L_{1M}, S_M-X-L_{1M}, and $Others$. That is, when an addressee continues or begins a mutual gaze with the other listener or looks at both speaker and the other listener with mutual gazes, the frequency of turn-taking is high.
- Addressees' gaze transition pattern has significantly high frequency at the time of S_M. That is, when an addressee continues a mutual gaze with a speaker, the frequency of turn-keeping is high.

Fig. 5 shows the frequencies of side-participants' gaze transition patterns under turn-taking and turn-keeping conditions using 4209 data. As a result, there are fifty-four gaze transition patterns. The "Others" class includes forty-one patterns, each of which are occurred in less than 1% of the data because the number of data is small. The result of a chi-squared test shows that the frequencies of gaze transition patterns differ significantly between turn-taking and turn-keeping ($\chi = 225.16$, $df = 13$, $p < .01$). Next, to verify which gaze transition patterns differ between conditions, we conduct a residual analysis [16]. The result is shown in 5, from which we understand the following.

- Side-participants' gaze transition patterns have significantly high frequencies at the time of L_1, L_{1M}, S-X-L_1, X-L_1, X-S-X-L_1, L_1-X-L_1, and $Others$. That is, when a side-participant begins looking at the other listener, continues looking at the other listener, looks at both a speaker and the other listener, or continue a mutual gaze with the other listener, the frequency of turn-taking is high.
- Side-participants' gaze transition patterns have significantly high frequencies at the time of S, X-S, and S-X-S. That is, when a side-participant continues or begins looking at a speaker or looks at a speaker several times, the frequency of turn-keeping is high.

Therefore, these results suggest that speaker's, addressee's, and side-participant's gaze transition pattern is valuable information for predicting turn-taking.

D. Proposed Turn-taking Prediction Model

To predict whether turn-taking or turn-keeping occurs, we propose a probabilistic model using the occurrence probability (frequency) of speaker's, addressee's, and side-participant's gaze transition patterns in turn-taking and turn-keeping situations in Subsection C of Section IV . We assume that each participant's gaze transition patterns are independent events and use a Nave Bayesian model such that:

$$P(y \mid f_{sp}, f_{ad}, f_{si}) \propto P(y) \cdot P(f_{sp} \mid y) \cdot P(f_{ad} \mid y) \cdot P(f_{si} \mid y). \quad (1)$$

$y = 1$ indicates the turn-taking and $y = 0$ indicates the turn-keeping. f_{sp} indicates speaker's gaze transition pattern, f_{ad} indicates addressee's gaze transition pattern, and f_{si} indicates side-participant transition pattern. $P(y = 1)$ indicates the occurrence probability of turn-taking and $P(y = 1)$ indicates the occurrence probability of turn-keeping. $P(f_{sp} \mid y = 1)$ indicates the conditional probability of speaker's gaze transition pattern in a turn-taking situation, $P(f_{sp} \mid y = 0)$ indicates the conditional probability of speaker's gaze transition pattern in a turn-keeping situation, $P(f_{ad} \mid y = 1)$ indicates the conditional probability of addressee's gaze transition pattern in a turn-taking situation, $P(f_{ad} \mid y = 0)$ indicates the conditional probability of addressee's gaze transition pattern in a turn-keeping situation, $P(f_{si} \mid y = 1)$ indicates the conditional probability of side-participant's gaze transition pattern in a turn-taking situation, and $P(f_{si} \mid y = 0)$ indicates the conditional probability of

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

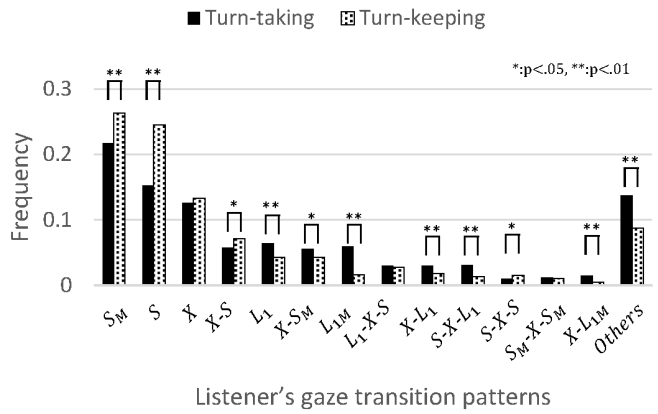

Fig. 6. Relationship between listener's gaze transition patterns and turn-taking

side-participant's gaze transition pattern in a turn-keeping situation. From formula (1), if $P(y = 1 \mid f_{sp}, f_{ad}, f_{si})$ is larger than $P(y = 0 \mid f_{sp}, f_{ad}, f_{si})$, the result of the prediction will be turn-taking. Otherwise, it will be turn-keeping.

If a speaker looks at places other than participants, we regard both listeners' roles as side-participants. Therefore, we use the following model instead of formula (1).

$$P(y \mid f_{sp}, f_{si}) \propto P(y) \cdot P(f_{sp} \mid y) \cdot \prod_{i=1}^{2} P(f_{si_i} \mid y). \quad (2)$$

f_{si_i} indicates side-participant si_i $(i \in 1, 2)$'s gaze transition pattern, $P(f_{si_i} \mid y = 1)$ indicates the conditional probability of listener si_i's gaze transition pattern in a turn-taking situation, and $P(f_{si_i} \mid y = 0)$ indicates the conditional probability of side-participant si_i's gaze transition pattern in a turn-keeping situation. In the case of formula (2) as well as formula (1), if $P(y = 1 \mid f_{sp}, f_{si_1}, f_{si_2})$ is larger than $P(y = 0 \mid f_{sp}, f_{si_1}, f_{si_2})$, the result of the prediction will be turn-taking. Otherwise, it will be turn-keeping.

V. Evaluation of turn-taking prediction

In this section, we evaluate the performance of the proposed model in Subsection D of Section IV. We use Ishii's prediction model as a base model, which uses gaze transition patterns without taking into account participation status [12]. The details of the models are as follows:

- Base model: The base model uses gaze transition patterns not taking into account participation status. In other words, this model does not distinguish between addressee and side-participant. The model is shown below.

$$P(y \mid f_{sp}, f_{l_1}, f_{l_2}) \propto P(y) \cdot P(f_{sp} \mid y) \cdot \prod_{i=1}^{2} P(f_{l_i} \mid y). \quad (3)$$

f_{l_i} indicates listener l_i $(i \in 1, 2)$'s gaze transition pattern, $P(f_{l_i} \mid y = 1)$ indicates the conditional probability of listener l_i's gaze transition pattern in a turn-taking situation, and $P(f_{l_i} \mid y = 0)$ indicates the conditional probability of listener l_i's gaze transition pattern in a turn-keeping situation. The occurrence probabilities of each twelve speaker's gaze transition patterns in turn-taking and turn-keeping situations (see Fig. 3) are used for conditional probability $P(f_s \mid y)$. In order to decide listeners' patterns used in base model, we analyze the relationship between listeners' gaze transition patterns and turn-taking in the same method as Subsection C of Section IV. Fig. 6 shows the result of analysis. Based a result of analysis, the the occurrence probabilities of each fourteen listener's gaze transition patterns in turn-taking and turn-keeping are used for conditional probability $P(f_{l_i} \mid y)$.
- Proposed Model: Our proposal model in Subsection D of Section IV is used. The occurrence probabilities of each eleven speaker's gaze transition patterns in turn-taking and turn-keeping situations (see Fig. 3) are used for conditional probability $P(f_{sp} \mid y)$. The occurrence probabilities of each fourteen addressee's gaze transition patterns in turn-taking and turn-keeping situations (see Fig. 4) are used for conditional probability $P(f_{ad} \mid y)$. The occurrence probabilities of each fourteen side-participant's gaze transition patterns in turn-taking and turn-keeping situations (see Fig. 5) are used for conditional probability $P(f_{si} \mid y)$.

TABLE I
Evaluation result of prediction model of turn-taking

Model	Precision	Recall	F-measure
Base model	0.506	0.415	0.451
Proposed model	0.517	0.439	0.470

In each model, $P(y = 1) = 0.613$ and $P(y = 0) = 0.387$, which are ratios of the number of data (the number of turn-taking is 2300 and turn-keeping is 1399), are given as prior probability.

We employ twelve-fold cross validation. 3619 data are divided into twelve sessions for each group of participants. One session is tested using the model trained with the other eleven sessions and this process is repeated twelve times by changing the training and testing sessions. Then, we calculate the average of estimation accuracy. The results of the evaluation are shown in Table I. The F-measure for the proposed model is 0.470, while that for the base model is 0.451. This suggests that dividing the listener's role into addressee and side-participant contribute to prediction turn-taking or turn-keeping.

VI. Discussion

First, we consider the reason why gaze transition patterns are different between addressee and side-participant. As can be seen in Fig. 2, the addressee tends to look at the speaker. This indicates that one recognizes that the speaker directs utterance to oneself and confirms that one is an addressee. In addition, one tells the speaker that one recognized oneself as the addressee by looking at the speaker. On the other hand, the side-participant tends to look at the other listener or both participants, or not to look at participants at all. This indicates

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

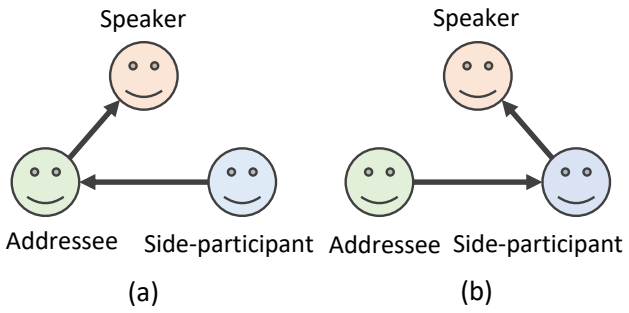

Fig. 7. Example where prediction varies depending on the role

that one confirms that the other listener is selected as the addressee by the speaker. In addition, one avoids becoming the addressee by not looking at participants or looking at the other listener. Lerner [6] mentions that there are conditions that the addressee notices the gaze of the speaker and the side-participant also have to know that the other listener is the addressee in order for the speaker to succeed in selecting the next speaker by gaze behaviors. This study also suggests that gaze behaviors are different between addressee and side-participant.

Next, we consider the reason why the precision of turn-taking prediction is improved by dividing listeners' roles into addressee and side-participant. In Figs. 5 and 6, for S in turn-keeping, the frequency of a side-participant is higher than that of a listener. In contrast, in Figs. 4 and 6, also for S in turn-keeping, the frequency of an addressee is far smaller than that of a listener. As we can see, we can actually find the two opposite the frequencies of gaze transition patterns in those that have been distinguished uniformly as a listener

Let us illustrate these situations in Fig. 7. In Fig. 7, (a) shows that the addressee gazes at the speaker (pattern S) and the side-participant gazes at the addressee (pattern L_1) and (b) shows that the addressee gazes at the side-participant (pattern L_1) and the side-participant gazes at the speaker (pattern S). In the conventional method, the probability of turn-keeping is higher than that of turn-taking in either situation (a) or (b). The model predicts turn-keeping. On the other hand, the proposed method predicts a result depending on the situation of (a) or (b) in Fig. 7. In (a), the probability of turn-taking is higher than that of turn-keeping. But, in (b), the probability of turn-keeping is higher than that of turn-taking.

We suppose that we are able to identify the similar opposite gaze transition pattern in Figs. 4, 5, and 6 although we do not indicate individual cases here. Therefore, it is possible that the proposed model can deal with the cases where base model cannot predict turn-taking correctly.

VII. Conclusion

In this paper, we have focused on participation status defined Goffman [4] in multi-party conversation and demonstrated the difference in gaze behaviors of each listeners' role, addressee and side-participant.

We find the difference of gaze transition patterns between addressee and side-participant. Based on results of analysis, we construct the probabilistic prediction model using the occurrence probabilities of gaze transition patterns for each listeners' roles. The result shows that the F-measure of the proposed model is higher that of the conventional method that does not consider participation status.

Although we define addressee as the person seen most frequently by the speaker near the end of an utterance, we cannot discuss whether method of accounting for listener's role is appropriate or not. Therefore, we need to conduct various judging methods of listener's role such as regarding the person who is seen last by the speaker as the addressee.

Several studies perform not only turn-taking prediction but also the next speaker prediction [12], [17]. Moreover, Ishii also predict the start timing of the next utterance. But, we can only predict turn-taking taking account of participation status. In the future, we plan to construct models to predict the next speaker and the timing of start timing of the next utterance considering participation status.

References

[1] A. Dielmann, G. Garau, and H. Bourlard, "Floor holder detection and end of speaker turn prediction in meetings," In INTERSPEECH, pp.2306-2309, 2010.
[2] A. Kendon. "Some functions of gaze direction in social interaction," Acta Psychologica, Vol. 26, pp.22-63, 1967.
[3] D. Schlangen, "From reaction to prediction experiments with computational models of turn-taking," In INTERSPEECH, pp.17-21, 2006.
[4] E. Goffman, " Forms of Talk," University of Pennsylvania, 1981.
[5] G. A. Levow, "Turn-taking in mandarin dialogue: Interactions of tones and intonation," In SIGHAN, 2005.
[6] G. H. Lerner, "Selecting next speaker: The context-sensitive operation of a context-free organization," Lannguage in Society, Vol. 32, pp. 177-201, 2003.
[7] H. H. Clark, "Using Language," Cambridge University Press, 1996.
[8] H. Sacks, E. A. Schegloff, and G. A. Jefferson, "A simplest systematic for organization of turn-taking for conversation," Language, Vol. 50, No. 4, pp. 696-735, 1974.
[9] K. Jokinen, K. Harada, M. Nishida, and S. Yamamoto. "Turn-alignment using eye-gaze and speech in conversational interaction." In INTERSPEECH, pp.2018-2021, 2011.
[10] L. Ferrer, E. Shriberg, ad A. Stolcke, "Is the speaker done yet? faster and more accurate end-of-utterance detection using prosody in human-computer dialog." In iInternational Conference on Spoken Language Processing, Vol. 3, pp. 2061-2064, 2002.
[11] I. de Kok and D. Heylen, "Multimodal end-of-turn prediction in multi-party meetings, " In ACM International Conference On Multimodal Interaction, pp.91-98, 2009.
[12] R. Ishii, K. Otsuka, S. Kumano, M. Matsuda, and J. Yamato, " Predicting next speaker and timing from gaze transition patterns in multi-party meetings," In ACM International Conference On Multimodal Interaction, pp.79-86, 2013
[13] R. Ishii, K. Otsuka, S. Kumano, M. Matsuda, and J. Yamato, "Using Respiration of Who Will be the Next Speaker and When in Multiparty Meetings," In ACM Transactions on Interactive Intelligent Systems, Vol. 6, No. 4, 2016.
[14] S. Duncan, "Some signals and rules for taking speaking turns in conversations," Journal of Personality and Social Psychology, Vol. 23, pp. 283-292, 1972.
[15] S. Duncan and D. W. Fiske, "Face-to-face interaction: Research, methods and theory," Hillsdale, NJ: Lawrence Erlbaum, 1977.
[16] S. J. Haberman, "The analysis of residuals in cross-classified tables," In Biometrics, Vol. 29, pp. 205-220, 1973.
[17] T. Kawahara, T. Iwatate, and K. Takanashi, "Prediction of turn-taking by combining prosodic and eye-gaze information in poster conversations," In INTERSPEECH, pp.9-13, 2012.
[18] Y. Den and M. Enomoto, "A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation. In Nishida, T. (Ed)," Conversational informatics: An engineering approach, pp. 3-7-330, Hoboken, NJ: John Wiley and Sons, 2007.
[19] Y. Den, H. Koiso, T. Maruyama, K Maekawa, K. Takanashi, M. Enomoto, and N. Yoshida, "Two-level annotation of utterance-units in Japanese dialogs: An empirically emerged scheme," International Conference on Language Resources and Evaluation, pp. 1483-1486, Valletta, Malta, 2010.

ISBN: 978-988-14049-0-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)