Eye-gaze Experiments for Conversation Monitoring
Kristiina Jokinen\(^1\) Masafumi Nishida Seiichi Yamamoto
Doshisha University
Department of Information Systems Design
1-3 Miyakodani, Tatara, Kyoutanabe-shi,
Kyoto 610-0321, Japan
+81-774-65-6983
[kjokinen | mnishida | seyamamo]@mail.doshisha.ac.jp

ABSTRACT
Eye-tracking technology has recently been matured so that its use in studies dealing with unobtrusive and natural user experiments has become easier to conduct. Simultaneously, human computer interactions have become more conversational in style, and more challenging in that they require various human conversational strategies, such as giving feedback and managing turn-taking. In this paper, we focus on eye-gaze in order to investigate turn taking signals and conversation monitoring in naturally occurring dialogues. We seek to build models that deal with the important aspects of which interlocutor the speaker is talking to, and what kind of turn taking signals the partners elicit, and we report the first results of our eye-tracking experiments.

Categories and Subject Descriptors

General Terms
Design, Experimentation, Human Factors, Languages

Keywords
Eye-tracking, human-human interaction, multiparty conversation.

1. INTRODUCTION
In human computer interaction, one of the main goals is to design systems that are natural and unobtrusive, yet robust and efficient. In particular, the prevailing ubiquitous computing paradigm assumes that our environment will be populated with several context-aware devices that communicate with each other, and are also capable of real-time robust interaction with users. Such applications range from various home appliances to robotic companions and, from the users’ point of view, it is expected that they accept input in speech and other modalities that best suit to the users’ needs and to the task in hand. This requires that the applications have appropriate communicative skills so that the users can interact with them in a natural and smooth manner. Moreover, as the devices can also request the user’s attention to a particular task or item in the environment, such skills must also include simultaneous coordination of action and communication.

Such communicative skills cannot be achieved only by improving interaction technology but also by advancing the system’s communicative capability [1], i.e. by modelling the users’ natural dialogue strategies and integrating the models in the system’s behavioural component. Dialogue strategies concerning feedback of accepted requests, acknowledgement of understood commands, management of turn taking etc. are often coordinated using non-verbal signals, and the systems should thus accept a wide variety of modalities in which the users convey meanings and control the interaction: they should be capable of sensing and interpreting communicative signals expressed by gestures, facial expressions, eye-gazing, and body posture.

In this paper, we focus on eye-gaze and its functions in managing turn-taking in natural dialogues. The importance of eye-gaze in order to coordinate and control interactive situations has been established by previous research, and the dimension of gazing most frequently investigated is the gaze direction which serves the functions of framing of interaction and establishing who is going to speak to whom and about what. Other aspects of gaze behaviour, such as how long and how often the partner is looked at, are also relevant from the communicative point of view, as well as the whole eye region (eyebrows, eyelids, wrinkles), and the upper body and gesture movements.

We pursue a multilevel analysis of the conversational data so as to integrate view-points from both signal processing and human communication studies. By studying human-human interactions we communicate intuitive human behaviour patterns that can also be deployed in human-computer interactions. Models concerning dependencies among different modalities and their use in the coordination of communication can then be used to enhance interaction possibilities for context-aware and intelligent devices.

The paper is structured as follows. We first provide a short overview of the eye-gaze and interaction research in Section 2. Our multilevel approach and the experimental setup are described in Section 3, and the results are discussed in Section 4. Finally, Section 5 draws conclusions and points to future research.
2. EYE-GAZE AND INTERACTION

The crucial role of eye-gazing in fluent communication has long been acknowledged in human-human interactions ([2], [3]). Eye-gazing regulates reciprocal arrangement and understanding in multi-party settings and is important in managing the flow of interaction. For instance, gaze provides rich and articulated feedback [4], and is important in establishing and sustaining the focus of shared attention [5]. Gaze helps to monitor behaviour when an object/person enters the field of view, or a person wants to interrupt and join the conversation. [6] shows how eye-gazing is used in signalling turn-taking in social interactions, while [7] gives evidence of the importance of eye contact for turn-taking in video-conferencing.

Eye-tracking technology has long been used in studies within ergonomics and cognitive psychology, but its use in interactive communication has only recently become an object of research. In particular, the techniques to measure and analyse eye movement has matured to the level which allows experimental settings to be easier (and cheaper) to arrange. Interface technology is actively growing, and e.g. eye-typing interfaces [8] and various Attentive User Interfaces [9] are being developed. Active research is also going on for modelling interactive behaviour for various artificial agents [10]. For instance, [11] describe an eye-gaze model for believable virtual humans and [12] built an eye-gaze model to ground information in interactions with embodied conversational agents. [13] demonstrates gaze modelling for conversational engagement and [14] considers the use of non-verbal behaviour, including eye-gaze, in a tourist-information setting.

The topic of the studies reported in this paper deals with the role of gaze as a means of controlling and coordinating interaction. Compared with the previous research, our studies differ in two respects: we focus on three-party conversations instead of two-party dialogues, and we are also interested in the role of eye-gaze in communication management in free-flowing conversations rather than the use of eye-gaze in task related dialogues.

3. EXPERIMENTS

As discussed above, gaze has many functions in face-to-face interaction. We investigate the following two hypotheses:

(a) eye-gaze is an effective means to coordinate turn-taking and the organization of talk

(b) eye-gaze shows the speaker’s intention to turn-yielding, continuing to talk (turn-holding) and preparing to talk

3.1 Method

Our method is hybrid in that it combines the top-down human approach with the signal analysis. Based on the findings in linguistics and cognitive science, we have selected features that are related to feedback and turn-taking, and track these in the data, and we model their importance and interdependence on the basis of their frequency and their temporal correlations present in the data. The multilevel analysis will provide a more comprehensive view of the dialogue and communication strategies than the analysis on a single level only, as we can simultaneously focus on both the use and perception of non-verbal information.

The data is analysed at the signal level and at the pragmatic dialogue level. On the signal level, we use the standard measures for speech analysis, and for eye-tracking (such as the participants’ gaze fixation and the gaze path between the partners), while on the dialogue level, we produce a general spoken dialogue analysis by annotating the data with important dialogue aspects observed in the speakers’ dialogue actions. Such aspects include facial expressions, turn management, feedback type (backchannels, acknowledgement, agreement, feedback elicitation), gesturing and body posture. Gaze events are thus analyzed in terms of their observable properties (length of fixations, direction, etc.), and by taking into account the context of interaction within which the gaze behaviour is interpreted according to pragmatic principles.

3.2 Setup

We collected data from speakers participating in natural dialogues. The collection setup is shown in Figure 1. Three participants sit in a triangle formation, and one of them (we call this person ES) has the eye-tracker in front of them to record eye movements (the rightmost person in Figure 1). The two other participants, the left-hand speaker (LS) and the right-hand speaker (RS) are videotaped so as to provide a reference to what ES sees and is looking at, and where ES’s gaze is focused on.

![Figure 1 Data collection setup.](image)

Six participants were recruited among the students. One of the participants was a female student but otherwise we did not aim at gender balance. They were all Japanese, and they knew each other but not necessarily very well. To get a mixture of participants with minimum contact with each other in the experimental setting, the participants rotated among themselves so that the eye-tracked person was always a new participant in each triad. The task of the participants was to learn more about the others and discuss issues that they were interested in. Consequently, the dialogues are natural chatting on topics that range from hobbies and weekend plans to studies and travelling.

We collected 6 dialogues each about 10 minutes long. For the annotation we chose about 4-6 minutes stretches from the beginning of each dialogue, so that the selected parts dealt with a cohesive topic. Three annotators annotated the files with dialogue acts, eye-gazing, facial expressions, hand gestures, and body posture. (However, since gestures and body posture by the eye-tracked participant was not seen on the video, these aspects were left out from the present study.) The dialogue act annotation followed the AMI corpus guidelines [15], while the non-verbal communication was annotated according to the modified MUMIN scheme [16]. The features included face, gaze direction,
whether gazing to the current speaker, head movement, feedback, turn taking, and emotion. The annotator agreement was measured by Cohen’s kappa-coefficient which calculates agreement beyond chance by distinguishing the observed agreement ($A_{obs}$) from the agreement by chance ($A_{ch}$), according to the following formula:

$$K = \frac{A_{obs} - A_{ch}}{1 - A_{ch}}$$

We reached the kappa value 0.46, which corresponds to moderate agreement according to the scale proposed by [17]. It must be noted that low kappa scores do not necessarily mean low agreement: if the annotators share certain assumptions of the data, their chance agreement is higher, and the above formula gives smaller kappa values. The annotations were further checked and agreed upon manually for the experiments. A view of the Anvil annotation board is shown in Figure 1.

**Figure 1 Anvil annotation board.**

Some basic statistics of the data is given in Table 1. Concerning the eye-tracked person’s gaze shifts, the data is not always clear as the gaze trace sometimes breaks. If the break (= no gaze) is shorter than 0.2 seconds, the elements were regarded as part of the same gaze event (unless there was a shift), otherwise they were regarded as separate events (and no shift could be recorded). The difference in face elements and gaze shifts is due to the annotation since face track elements also contain head movement and sometimes the gaze does not shift during head movement.

**Table 1 Basic statistics of the annotated dialogue samples.**

<table>
<thead>
<tr>
<th></th>
<th>Eye-tracked speaker</th>
<th>Left speaker</th>
<th>Right speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialogue acts</td>
<td>146</td>
<td>118</td>
<td>101</td>
</tr>
<tr>
<td>Total speaking time</td>
<td>3:39,15</td>
<td>3:23,37</td>
<td>2:21,58</td>
</tr>
<tr>
<td>Ave length of dact /s</td>
<td>1.50</td>
<td>1.72</td>
<td>1.40</td>
</tr>
<tr>
<td>Max /s</td>
<td>7.47</td>
<td>8.61</td>
<td>13.31</td>
</tr>
<tr>
<td>Min /s</td>
<td>0.03</td>
<td>0.33</td>
<td>0.50</td>
</tr>
<tr>
<td>Face elements</td>
<td>181</td>
<td>182</td>
<td>273</td>
</tr>
<tr>
<td>Ave length of gaze /s</td>
<td>1.67</td>
<td>2.91</td>
<td>1.94</td>
</tr>
<tr>
<td>Max /s</td>
<td>11.6</td>
<td>11.48</td>
<td>10.11</td>
</tr>
<tr>
<td>Min /s</td>
<td>0.03</td>
<td>0.33</td>
<td>0.17</td>
</tr>
<tr>
<td>Gaze shifts</td>
<td>100</td>
<td>90</td>
<td>143</td>
</tr>
</tbody>
</table>

### 4. RESULTS

#### 4.1 Turn-taking prediction

Evidence of gaze both as turn yielding and turn holding cues came first from [3] who observed that utterances which terminated without a speaker gaze had delayed listener responses more frequently than those with a mutual gaze. Also [7] reports about videoconference settings that the lack of gaze can decrease the interlocutors’ turn-taking efficiency by 25%.

We investigated if turn-taking could be predicted on the basis of the gaze elements and the features that were annotated in our data. Predicting the turn category, i.e. hold vs. shift, is a two-class classification task, and we used the Support Vector Machine technique in the Weka software package [18] for this. Using the seven features (with the number of feature values ranging between 3 and 7), we classified the gaze instances with respect to turn management, and received weighted average precision of 0.663, recall 0.676, and f-measure of 0.665. If we ignoring the gaze information from the feature set, the classification results were slightly worse, see Table 2. As a baseline classification, we used Weka’s ZeroR algorithm which selects the most common category.

Comparing the SVM figures with each other and with the baseline, we can see that the classification with gaze features improves both recall and accuracy, although the difference to classification without the gaze is not huge. As the data set is rather small, we, of course, have to be careful when drawing conclusions. It seems, however, that the data and the annotated features confirm the earlier findings, namely that eye-gazing is important in turn management, but they also provide support for the new hypothesis that there are other cues as well which are crucial in signalling turn taking. As the difference between with-gaze and without-gaze classifications is not big, our data begs for this kind of an explanation. In particular, the other cues may include head movement and facial expressions, and in the next sections we discuss their role in turn management more closely.

**Table 2 Results for turn-taking prediction.**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Precision</th>
<th>Recall</th>
<th>f-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM with gaze</td>
<td>0.663</td>
<td>0.676</td>
<td>0.665</td>
</tr>
<tr>
<td>SVM without gaze</td>
<td>0.651</td>
<td>0.672</td>
<td>0.645</td>
</tr>
<tr>
<td>ZeroR</td>
<td>0.420</td>
<td>0.648</td>
<td>0.510</td>
</tr>
</tbody>
</table>

#### 4.2 Turn taking signalling

It has been generally observed ([6], [17]) that gaze is important in signalling the partner the end of the turn and showing focus of attention. Turn-yielding cues are used by participants in deciding when they should start to speak: such cues usually include for instance a drop in pitch and/or loudness towards the end of the utterance, an unfilled pause after the completion of a topic sequence, termination of gesture, the speaker gazing back up to an interlocutor, and a shift of posture such as leaning back. As for the turn-holding signals, gaze aversion and gesticulation seem to be the ones commonly used.

Based on the hypothesis outlined at the end of Section 4.1, we wanted to investigate the effects of the communicative signals,
i.e. the head movement, perceived emotion, and gazing upon turn management, and as the first step, we ranked the attributes according to their contribution to turn management. For this, we evaluated each attribute with respect to the information gained by selecting it for turn management, and ranked the results with the Ranker method in Weka, using 10-fold cross-validation. The most highly ranked attributes are shown in Table 3.

As can be seen the best indicators for turn management are related to head movement and the perceived emotional state rather than eye-gaze. This is somewhat surprising but can be understood in the context of multiparty conversations: the interlocutors’ head movements may be more noticeable signals to the partners than eye-gazing, and thus they can also function as more reliable cues for turn management. The interlocutors effectively turn their head to the speaker if they want to take the turn, and if they do not want to take the turn, they move their head towards the other partner who will talk next.

In our data, the participants use gaze as a turn-taking signal but apparently it is not as important a signal as the head movement. This seems reasonable if we consider the fact that gaze is related to indicating one’s focus of attention. The interlocutors often use quick eye-gaze shifts to check their partners’ “current state of attention”, i.e. to learn if the current speaker is getting ready to yield the turn, or to elicit information whether the partners are still following one’s presentation. However, an interlocutor will be the next speaker only if his gaze is reciprocated by the current speaker. If the current speaker looks at the other partner, it is likely that this partner will be the next speaker as he already is at the current speaker’s focus of attention and thus under social obligation to continue. Moreover, as the speaker is also likely to have the head turned towards him, he has a better opportunity to notice when the speaker wishes to yield the turn, and can thus quickly grab the turn if necessary. On the other hand, if the interlocutor does not want to speak next, they can turn their head towards the other partner to indicate that their attention is directed to the partner who is expected to speak next.

Manual inspection of the data revealed that the speakers usually turn their head first and shift their gaze about one frame, or 30 milliseconds later, as if aligning their focus of attention. However, an interlocutor will be the next speaker only if his gaze is reciprocated by the current speaker. If the current speaker looks at the other partner, it is likely that this partner will be the next speaker as he already is at the current speaker’s focus of attention and thus under social obligation to continue. Moreover, as the speaker is also likely to have the head turned towards him, he has a better opportunity to notice when the speaker wishes to yield the turn, and can thus quickly grab the turn if necessary. On the other hand, if the interlocutor does not want to speak next, they can turn their head towards the other partner to indicate that their attention is directed to the partner who is expected to speak next.

To study the interdependence between the speakers’ gaze shifts and turn management, we analysed the data with respect to turn-taking and gaze-shifting possibilities. We calculated the occurrence of gaze shifts for each interlocutor depending on who is the current speaker and what is the speaker’s turn management strategy, i.e. if the speaker takes, holds, or gives the turn. The left and right speaker’s shifts were combined but the eye-tracked speaker’s gaze shifts are given separately for them, the eye-gaze events are accurate due to the eye-tracker but the other features, facial expressions and head movements, are missing due to the video positioning.

The participants’ percentage number of eye gaze shift is given in Table 4. As can be seen, the shift mainly depends on who is speaking: the current speaker receives most eye gazes. In the case of LS and RS, the distribution is clear if we ignore the gaze shifts to other (i.e. neither of the participants). The eye-tracked person, when speaking, divides eye gaze between the left and right speaker roughly in an equal way, 57% and 43%, respectively, but when the other interlocutors speak, they overwhelmingly shift gaze to somewhere else (45%) or to the eye-tracked person (40%), and very little to the partner (15%), obviously because ES sits in front of them but the partner is located on the side.

Table 3 Ranking of attributes for turn taking

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Average merit</th>
<th>Average rank</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HeadMovement</td>
<td>0.079 +/- 0.012</td>
<td>1 +/- 0</td>
<td></td>
</tr>
<tr>
<td>HeadRepetition</td>
<td>0.052 +/- 0.01</td>
<td>2.2 +/- 0.4</td>
<td></td>
</tr>
<tr>
<td>Emotion</td>
<td>0.045 +/- 0.005</td>
<td>3 +/- 0.63</td>
<td></td>
</tr>
<tr>
<td>GazeInterlocutor</td>
<td>0.032 +/- 0.006</td>
<td>3.8 +/- 0.4</td>
<td></td>
</tr>
<tr>
<td>GazeDirection</td>
<td>0.019 +/- 0.005</td>
<td>5.4 +/- 0.49</td>
<td></td>
</tr>
<tr>
<td>Face</td>
<td>0.017 +/- 0.004</td>
<td>5.6 +/- 0.49</td>
<td></td>
</tr>
</tbody>
</table>

Table 4 Percentage number of gaze shifts of the eye-tracked speaker and the combined left and right speakers, given the current speaker (self, left, right, eye, partner), and total.

Table 5 shows a closer study on the interlocutors’ gaze shifts with respect to turn management, i.e. whether the current speaker takes the turn, holds the turn, or yields the turn. It reveals that ES always looks at the partner who takes the turn, and LS and RS do so mostly if the ES takes the turn, but if the partner takes the turn, they focus their attention on the turn-
taking partner only about 58% of the time, and otherwise on other objects. The table also shows nicely how the interlocutors tend to shift gaze somewhere else when the partner gives the turn, while the eye-tracked person seems to divide gaze shifts between the two partners.

Although the interlocutors seem to have similar overall gaze shift distributions over the current speaker and the speaker’s turn management, their gaze shifts differ depending on whether they are speaking or listening. In other words, gaze shifts during one's own turn management differ from gaze shifts when one observes the partners speaking. This sounds reasonable, since speaking and listening require different roles and actions in the conversation: the speaker needs to observe the partner’s reactions and possible attempts to speak, while the listener needs to observe the partner’s possible signals to stop speaking and find a suitable point to enter the conversation.

**Table 6 Gaze shifts of ES and combined LS&RS given one's own turn management.**

```
<table>
<thead>
<tr>
<th>Gaze</th>
<th>Other</th>
<th>ES</th>
<th>LS</th>
<th>RS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn Take</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyetracked turn management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn Hold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn Give</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner's turn management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Table 7 Gaze shifts of ES and combined LS&RS given the partner's turn management.**

```
<table>
<thead>
<tr>
<th>Gaze</th>
<th>Other</th>
<th>ES</th>
<th>LS</th>
<th>RS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn Take</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyetracked turn management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn Hold</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turn Give</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner's turn management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Tables 6 and 7 compare the participants’ behaviour as a speaker and a listener. There is a clear difference between the two cases, i.e. whether one speaks oneself or listens to other participants speaking. When the interlocutor takes the turn to speak, gaze is usually shifted to somewhere else (69%) than to either of the partners, whereas when one listens the partner to take the turn, gaze is usually directed to this partner, and only 19% of the time to somewhere else. When the interlocutor holds the turn, gaze is usually divided between the ES (sitting in front of the speaker) and other objects, i.e. neither LS or RS look at each other when they hold the turn. However, when the partner holds the turn, they divide their gaze rather evenly between the speaker, the ES and other object in the environment. Finally, looking at the turn giving events, if it is the partner who yields the turn, one generally looks away somewhere else and only about 20% of time looks at the partner, whereas if one yields the turn oneself, gaze is shifted to the partners and only about 22% of time to other objects. The results confirm the hypothesis that at the end of one's turn, the speaker looks at the partners in order to elicit feedback and possible turn taking, while at taking the turn, the speaker looks away and thus signals the acceptance of the turn as well as being focussed on the planning and production of the turn. During the course of the interaction, the gaze shifts seem to be towards the partners about 60% of the time and to other objects about 40% of the time.

Of course, these tendencies fit together well and also confirm the previous research results: when accepting the turn, the speaker usually looks away. On the other hand, the data also shows that when the partner signals yielding of the turn, the participant shifts gaze to something else so as not to accept the turn. The shifts while the partner is still speaking and finishing the turn are counted as partner turn-yielding gaze shifts, while the shifts after the speaker has already started own speech are turn-taking shifts.

The gaze shifts distributions may be partly explained by the setup: ES sits in front of LS and RS and thus received much of the eye-gazing, and also, being aware of the experiment, ES focussed his eyes on the partners rather than other targets. However, we can also assume some genuine conversational principles that underlie the gaze behaviour, such as control of the interaction and the implicit roles of the participants. For instance, ES seems to chair the interactions which warrants him to take more initiative (i.e. ask questions), but also obliges ES to share his attention with the two partners evenly, unlike the other partners who can more freely focus their attention to something else. However, these aspects need further research.

### 5. CONCLUSIONS

Recent developments in speech, eye-tracking, and gesture recognition technologies form enabling technologies which can greatly enhance future human-technology interfaces by allowing natural, intuitive, easy, and friendly interactions in ubiquitous environment. We believe that to achieve these goals, it is also important to deepen our knowledge of the interaction strategies and processes that underlie fluent and robust human-human communication. The multi-level analysis of interactive situations, as proposed in this paper, is a step towards this direction as it provides a more comprehensive view of dialogue phenomena than an analysis on a single level only. It can thus help us to reveal correlations and dependences among various modalities, and to build models which can prove useful in building intelligent interactive agents. The eye-gaze experiments reported in this paper thus aim to contribute to the theories of human-human interaction and to the modelling of gaze behaviour for the design and development of robust and natural interactions with smart objects, services, and environments.

Although our experiments are based on a small data set, they still provide some interesting observations and results concerning non-verbal cues in turn taking and turn-signalling, especially on the role of eye-gaze in turn management. In particular, we noticed that in multiparty conversations, the turn management seem to be signalled with head turns rather than eye-gazing, while eye-gaze is important as an initial signal of who could be next speaker. Eye-gaze also effectively allows the partners to reach agreement of the next speaker as part of the mutual
agreement between the speaker and the next speaker in their short mutual gazing. We also noticed that the eye-gaze shifts are different depending on the interlocutor's role in the dialogue activity: i.e. whether the interlocutor is speaking or listening to other partners to speak.

Some interesting challenges concern the correlations between eye-gaze, speech, and dialogue events due to their different temporal scales. We will investigate these further, and will also test the hypotheses in a larger corpus. In order to verify the role of eye-gaze in multiparty conversations, we will also specify the different features and their contributions to the prediction of turn-taking in a more detailed way.

Our future plans concern collecting and analysing more data, including conversations with participants who do not know each other in advance. This would allow us to compare eye-gazing behaviour with familiar and unfamiliar contexts. Moreover, we would like to include an analysis of gestures in the research, and produce some comparison of non-verbal behaviour across cultures as outlined in [19]. We also plan to include more thorough analysis of the speech features of the data so as to investigate the relationship between turn-taking and prosodic features in spontaneous dialogues. Finally, to visualise verbal and non-verbal communication and to spell out differences between the observed signals and the interpreted human communication elements, we plan to compare signal-level analysis of the video with the human annotated data, along the lines described in [20].

6. ACKNOWLEDGMENTS

We would like to thank all the students who helped us with the experiments and data annotation, and our subjects for taking part in the experiments. The first author would also like to thank NiCT for the grant that allows her to conduct research in Japan.

7. REFERENCES
