

Non-verbal signals for turn-taking and feedback

Kristiina Jokinen

University of Helsinki

Department of Speech Sciences

PO Box 9

00014 University of Helsinki

E-mail: Kristiina.Jokinen@helsinki.fi

Abstract

This paper concerns non-verbal communication, and describes especially the use of eye-gaze to signal turn-taking and feedback in conversational settings. Eye-gaze supports smooth interaction by providing signals that the interlocutors interpret with respect to such conversational functions as taking turns and giving feedback. New possibilities to study the effect of eye-gaze on the interlocutors' communicative behaviour have appeared with the eye-tracking technology which in the past years has matured to the level where its use to study naturally occurring dialogues have become easier and more reliable to conduct. It enables the tracking of eye-fixations and gaze-paths, and thus allows analysis of the person's turn-taking and feedback behaviour through the analysis of their focus of attention. In this paper, experiments on the interlocutors' non-verbal communication in conversational settings using the eye-tracker are reported, and results of classifying turn-taking using eye-gaze and gesture information are presented. Also the hybrid method that combines signal level analysis with human interpretation is discussed.

1. Introduction

Human conversations are surprisingly fluent concerning the interlocutors' turn-taking and feedback giving behaviour. Many studies have shown the accurate timing of utterances and pointed out that the speakers synchronize, or align their behaviour so as to provide robust and efficient communication. In the context of Interaction Technology, especially when considering applications like robotic companions which interact with the user in real time, such synchronization is also important: in order to allow smooth communication between an intelligent agent and the user, the agents need to have realistic models about when to take turns, when to interrupt, and how to catch the partner's attention. The agents have sensors that can detect the user's movements and enable recognition of communicatively relevant events, but it is also necessary to catch the user's attention if something important or unexpected has happened in the environment that needs to be communicated to the user. All this presupposes that the agents possess appropriate linguistic skills as well as models of turn-taking and feedback so as to successfully coordinate their action and communication, i.e. they need to act in a rational manner to synchronize their intentions (cf. Jokinen 2009a).

The focus of this paper is on eye-gaze and gesturing, and how they support coordination of interaction in natural dialogue situations. In particular, the paper deals with the eye-tracking technology and its use in the analysis of the turn-taking and feedback. The method proposed in the paper is hybrid in that it combines bottom-up signal-level analysis with top-down manual annotations: the approach uses novel interaction technology as well as human interpretation to analyse the interlocutors' communicative behaviour, and produces a multi-level analysis of the collected data on both signal and dialogue levels. Each significant communicative event is classified in terms of

its observable properties recognized automatically, as well as of its interpretation in the communicative context within which it occurs. Such multi-level analysis provides a more comprehensive view of the dialogue phenomena than an analysis on a single level only.

The structure of the paper is as follows. Section 2 gives an overview of the previous interaction research, while Section 3 presents data and the data collection setup, and Section 4 describes data annotation. Classification results concerning the relation between eye-gaze, gestures, and turn-taking are discussed in Section 5. The hybrid method, conclusions, and future work are discussed in Section 6.

2. Eye-gaze and interaction

The role of eye-gaze in fluent communication has long been acknowledged. It has been recognized as a relevant aspect of human-human interactions (Argyle and Cook 1976) as well as in developing understanding of shared attention with babies (Treverhatten, 1984). Kendon (1967) was one of the first to emphasize gaze as a turn yielding and turn holding cue: he observed that listener responses were quicker if there was a mutual gaze and that they were delayed if the previous utterance terminated without a speaker gaze. Recently eye-gaze has been discussed in reference to social interaction (Bavelas, 2005) and video-conferencing (Vertegaal 2003), providing further evidence for the importance of eye gaze in smooth turn-taking. In human-computer interaction, the research on embodied conversational agents and virtual humans has also used eye-gaze information to build more believable characters. For instance, Lee et al. (2007) describe an eye-gaze model for believable virtual humans, while Nakano and Nishida (2007) built an eye-gaze model to ground information in interactions with embodied conversational agents.

Eye-tracking is a process that records eye movements and allows one to determine where the person's gaze is fixed

at a given time. Although eye-trackers have long been used in human-computer interaction and cognitive psychology studies, their use in interaction research has only quite recently become more feasible, see e.g. Jacob and Karn (2003) for an overview of the technical evolution. Gaze-based interface technology, such as eye-typing interfaces (Majaranta and Räihä, 2002), has been actively developed, and eye-trackers have also been used to collect data for building interaction models. For instance, Ishii and Nakano (2008) used the eye-tracker to collect gaze data for a model that estimates the user's engagement in conversations based on their gaze behaviour, whereas Jokinen et al. (2009) describe experiments on using eye-tracker information to study turn-taking behaviour in natural conversations. Besides gaining insight into where the participants are looking at, eye-trackers thus allow us to add objective information into modelling. The research reported in the paper continues this kind of research and explores the use of eye-tracker information on the modelling of turn-taking and feedback functions in human-human communication.

3. Data

The data was collected at the Doshisha University in Japan. The collection and available data are described in Jokinen et al. (2010b) in more detail. The corpus contains 28 multiparty conversations, each about 10 min long, with three participants chatting about everyday things and activities that interest them. In each triad, the participants are either familiar or not familiar with each other, and there are also four conversations with female-all groups and with English-speaking participants. The task of the participants was to learn more about each other and they were encouraged to discuss issues that they were interested in. The conversations are thus natural chatting, and the topics deal with hobbies, weekend plans, studies, and travelling.

The basic setup is shown in Figure 1. Three participants sit in a triangle formation and the eye-tracked 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 and they provide a reference to where ES's gaze is focused on. A snapshot of eye-gaze data with a gaze path is shown in Figure 2.



Figure 2: Data collection setup.

Eye-gaze cannot always be recorded. This happens of course when ES blink their eyes, and in particular when

they laugh or if the ES eyes become so small that the relevant eye-patterns cannot be found. Compared with typical HCI eye-tracking experiments the situation is now also more complicated. Instead of ES looking at the stimulus on a stable computer screen, we have a group of three conversational partners who can move according to their conversational activity: they can tilt their head, gesture with their hands, bend their body forward, backward and sideways, etc. Although the optics of the eye-tracker is rather robust, it also has limitations as to the head movements by the ES, and thus ES was especially instructed not to make excessively large movements. It is possible that this had some effect on ES gesturing as they might have consciously tried to move less than normal, but the inspection of the data showed that in general this was not so: head movements of the partners were rather similar independently of whether the person appeared as ES or LS/RS. The reason may be the setup which supported ES looking straight ahead to the partners so that big sideways head movements were not necessary for ES to look at the partners' face. It was also suggested that because Japanese people in general gesture little, these kinds of technical constraints did not have any noticeable effect on the naturalness of the conversations. The low number of hand gestures was also attributed to similar cultural characteristics, but these aspects of course need to be substantiated with intercultural comparison studies (cf. discussion concerning future work in Section 6).



Figure 1: Camera view of the two participants showing a gaze path from left to right and a gaze fixation on the right eye of the right-hand side person.

4. Annotations

For the experiments reported in this paper we used six of the conversations among familiar participants and analysed about five minute clips of each, i.e. half of each conversation, all together about 30 minutes. The analysis was done at the signal level and at the dialogue level. On the signal level, information concerning the participants' gaze fixation and gaze path was recognized by the eye-tracker, while on the dialogue level, an overall spoken dialogue analysis was produced by manually annotating important dialogue features in the speakers' observed dialogue actions. The main goal of the annotation was to study the relation between non-verbal communication events such as eye-gaze, facial expressions, hand gestures, and body movement, and their communicative functions in turn-management and feedback giving processes. The annotations thus concern dialogue acts, gaze, and gestures,

as well as their communicative functions in terms of turn-management and feedback. As for the dialogue act annotation, the definitions developed for the AMI project (www.amiproject.org) were followed, whereas for the other annotation features, a modified MUMIN multimodal annotation scheme (Allwood et al., 2007) was applied. This included ES gaze path, coded with the feature *GazeObject* which refers to the object of ES's focus of attention. The value *NoGaze* refers to the time spans when there is no gaze, either because ES blinked, laughed, turned head badly, etc., which prevented the tracker from recording the gaze. If *NoGaze* is shorter than 0.2 seconds, the gaze elements were regarded as part of the same gaze event (unless there was a gaze shift), otherwise they were considered separate gaze events, but obviously no shift in between them could be recorded. The features and feature values are shown in Tables 1 and 2 for ES and LS/RS, respectively. The features differ slightly for the obvious reasons: the recordings did not include gaze paths for LS/RS but this is inferred from the participants' head and face features, while ES does not have facial display and hand gesture data (the annotation was done on the pilot data which was collected using only one camera – in the main data collection a second camera was also used and this recorded also ES). The different features are marked in italics in the tables.

	Annotation features	Feature values
Words	Dialogue Act	Backchannel, Stall, Fragment, BePositive, AskUnderstand, GiveUnderstand, AskAssesment, GiveAssesment, Suggest-offer, Inform, Ask, Other
	FeedbackDirection	Give, Elicit
	Feedback	CPU, Agree, NonAgree
	Turn	Give, Take, Hold, None
	Emotion/Attitude	Happy, Sad, Interested, Uninterested, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Disappointed, Satisfied, Other
	<i>GazeObject</i>	RS, LS, Other, NoGaze
	FeedbackDirection	Give, Elicit
	Feedback	CPU, Agree NonAgree
	Turn	Give, Take, Hold, None
	Emotion/Attitude	Happy, Sad, Interested, Uninterested, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Disappointed, Satisfied, Other
	<i>GazeObject</i>	RS, LS, Other, NoGaze
	FeedbackDirection	Give, Elicit
	Feedback	CPU, Agree NonAgree
	Turn	Give, Take, Hold, None
	Emotion/Attitude	Happy, Sad, Interested, Uninterested, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Disappointed, Satisfied, Other
	<i>GazeObject</i>	RS, LS, Other, NoGaze

Table 1: Annotation features for the eye-tracked person.

Annotation was carried out by three Japanese students using the Anvil software (Kipp, 2001). The annotators were requested to select communicative events that they considered important on the video, and annotate these

according to the annotation categories. They were then requested to compare their annotations, and select those events that they considered important and which had the start and end times within +/- 0.1s from each other. These events, which were observed and considered important by all the annotators, were used as input for the classification experiments later on.

	Annotation features	Feature values
Words	Dialogue Act	Backchannel, Stall, Fragment, BePositive, AskUnderstand, GiveUnderstand, AskAssesment, GiveAssesment, Suggest-offer, Inform, Ask, Other
	FeedbackDirection	Give, Elicit
	Feedback	CPU, Agree, NonAgree
	Turn	Give, Take, Hold, None
	Emotion/Attitude	Happy, Sad, Interested, Uninterested, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Disappointed, Satisfied, Other
	<i>Face</i>	Basic, Smile, Laughter, Scowl, Other
	<i>GazeDirection</i>	Up, Down, Side, Other
	<i>GazeInterlocutor</i>	toES-speaking, toES-notspeaking, toPartner-speaking, toPartner-notspeaking, awayFromInterlocutors
	<i>HeadMovement</i>	Nod, Jerk, Backward, Forward, Tilt, TurnToPartner, TurnSide, Waggle, Other
	<i>HeadRepetition</i>	Single, Repeated
	FeedbackDirection	Give, Elicit
	Feedback	CPU, Agree NonAgree
	Turn	Give, Take, Hold, None
	Emotion/Attitude	Happy, Sad, Interested, Uninterested, Surprised, Disgusted, Angry, Frightened, Certain, Uncertain, Disappointed, Satisfied, Other
	<i>Handedness</i>	RightHandForward, RightHandForward, RightHandSide, RightHandUp, RightHandDown, RightHandOther
	<i>Trajectory</i>	LeftHandForward, LeftHandForward, LeftHandSide, LeftHandUp, LeftHandDown, LeftHandOther
	<i>Hand-repetition</i>	Single, Repeated

Table 2: Annotation features for the left and right partners.

The annotations reached the kappa value 0.46 which corresponds to moderate agreement. It must be emphasized that a low kappa score does not necessarily mean low agreement. If the annotators share certain assumptions of the data, their chance agreement is higher and consequently kappa value is smaller. Figure 3 presents raw annotation statistics by two of the annotators, and shows that the numbers of the events that the annotators have selected for their annotation are rather equal. In other words, the annotators have independently interpreted many of the events occurring on the video as communicatively important. Manual comparison of the annotations shows that this is indeed the case: the annotations differ mainly on the segmentation of whether a candidate event is regarded as making up one or two elements, rather than whether the event as such has a communicative meaning. This gives reasons to assume that the annotators, who observe the conversational participants' behaviour from a certain view-point outside the actual dialogue activity, notice certain non-verbal communicative events in a fairly similar manner, although their interpretation of the communicative function of the particular event can widely differ.

EyeTracked.words: 47 elements	EyeTracked.words: 44 elements
EyeTracked.Gaze: 131 elements	EyeTracked.Gaze: 111 elements
SpeakerLeft.words: 38 elements	SpeakerLeft.words: 41 elements
SpeakerLeft.FacialDisplay: 61 elements	SpeakerLeft.FacialDisplay: 63 elements
SpeakerLeft.HandGesture: 6 elements	SpeakerLeft.HandGesture: 2 elements
SpeakerLeft.BodyPosture: 1 elements	SpeakerLeft.BodyPosture: ---
SpeakerRight.words: 53 elements	SpeakerRight.words: 59 elements
SpeakerRight.FacialDisplay: 52 elements	SpeakerRight.FacialDisplay: 49 elements
SpeakerRight.HandGesture: 24 elements	SpeakerRight.HandGesture: 11 elements
SpeakerRight.BodyPosture: 7 elements	SpeakerRight.BodyPosture: 4 elements

Figure 3: Comparison of basic statistics of the annotations by two annotators.

5. Experiments and results

Turn management refers to the regulation of the interaction flow in conversation with the goal of minimising overlapping speech and pauses. It is coded by the four general features in the annotation: *Turn Give*, *Turn Take* and *Turn Hold*. The value *TurnNone* refers to situations when the partner is listening and has no turn. In order to study how the annotation features characterize turn taking, classification of the events with respect to the four turn management possibilities was produced using the SVM (Support Vector Machine) algorithm. Table 3 presents the results (from Jokinen et al., 2010a).

Dataset	Majority	SVM
SVM LSRS with gaze	73.11 +/- 0.51	80.00 +/- 4.08
SVM LSRS without gaze	73.11 +/- 0.51	78.14 +/- 3.57
SVM ES with gaze	51.68 +/- 0.52	92.68 +/- 1.32
SVM ES without gaze	51.68 +/- 0.52	90.56 +/- 1.07

Table 3: The effect of gaze on turn-taking.

As can be seen, SVM improves on the majority

classification which always selects the most frequent category, but it is somewhat surprising that gaze does not have such an impact on the results as might have been expected. The classification seems to confirm the earlier findings that eye-gaze is an important signal for the coordination and control of dialogues but not a significant factor (Jokinen et al., 2009). If the speaker looks at the partner, it is likely that the partner will be the next speaker as he is already at the current speaker's focus of attention and thus under social obligation to continue. On the other hand, if the partner does not want to speak next, they can shift their gaze away so as to indicate that their attention is directed somewhere else, or they can turn their head towards another partner so that this partner becomes under obligation to speak next. In two-party dialogues, this works so that the speaker usually gazes the partner in order to elicit feedback concerning whether the partner still follows the presentation, and the listener gazes the speaker so as to give feedback whether they focus on the current speaker or wish to take the turn themselves. In multiparty dialogues, however, turning one's head towards the speaker seems a better signal for turn-taking, since the sudden head movement is a more visible and "bold" attention catcher than just gazing at the speaker (unless the staring is unexceptionally intense and lasts long). The interlocutors usually gaze at their partners in a quick and repeated manner in order to check the partners' "state of attention", but if the gaze is accompanied by a head turn, the action becomes more prominent and is likely to be interpreted by the partners as an intention to take or yield the turn. There is thus a better opportunity to be noticed and to control the conversation, and also to make it clear to the other participants that one is focussed on taking the turn.

Besides head movement, it can also be thought that hand gestures function in a similar manner as attention catchers, and are used to emphasise and make the interlocutors' communicative intentions clear (cf. raising one's index finger or the whole hand when requesting a turn in more formal situations). Since the data also has hand gestures annotated, these features were used to produce analogous classification of turn taking with respect to gesture information. However, the results turn out to be similar to those with gaze information. They are given in Table 4. SVM seems to improve mere majority classification, but the difference between classifications using vs. not using gesture information is not significant.

Dataset	Majority	SVM
SVM LSRS with gesture	48.8 +/- 0.01	79 +/- 0.04
SVM LSRS without gesture	47.1 +/- 0.01	74 +/- 0.04
SVM ES with gesture	67.6 +/- 0.01	90 +/- 0.01
SVM ES without gesture	65.0 +/- 0.01	87 +/- 0.01

Table 4: The effect of gesture on turn-taking.

This result is somewhat unexpected too, but it may be due to a small data set, or that the hand gestures in the particular data set are used for other functions than the

coordination of turn-taking. Firm conclusions would require further analysis and experimentation. However, it can also be hypothesised that non-verbal signals do not provide definite information about the interlocutor's turn-taking intentions as such, but need to be interpreted with respect to other non-verbal and verbal signals available in the communicative context. It may be that the repertoire of various signals involving gaze or hand is vague or too ambiguous in order to convey turn-taking information clearly, but it seems more reasonable to conclude that non-verbal signals simply have no conventional meaning associated with them that could be pinned down to their possible communicative functions. Using semiotic terminology, non-verbal signals are not symbols like words and utterances, but indexical signs that have certain relationship to the communicative situation they occur in. The relation is not conventionally agreed among the members of the speech community, however, but learnt through general causal laws and one's own activity in the environment. Naturally occurring non-verbal signals thus form patterns which must be interpreted against the whole communicative background rather than regarded as independently defined communicative symbols that the interlocutors exchange in order to reach some intended effect (e.g. turn taking). Following this line of reasoning, we can assume that the difference between head movement and hand gesturing in signalling turn taking is related to their indexical properties in communicative situations. If we assume that successful turn-taking requires mutual gaze and that head movements have strong relation to the changes in the speaker's eye-gaze, it is possible to conclude that head turns form strong indexical relations to turn-taking since they can indicate that the speaker is preparing for a mutual gaze. Hand-gestures, however, only attract the partner's attention but leave the mutual gaze open, so their relation to turn-taking is often less clear.

This view of non-verbal communication also accords with what was discussed above concerning the annotation method. The annotators were expected to select important and visible non-verbal events if these play an observable communicative function, and indeed, they selected much of the same events, i.e. human perception seems rather uniform in picking up similar events as important conversational phenomena. However, interpretations assigned to the events differed, and they also depended on how detailed gesturing the annotators considered as forming a single communicative message. The reasoning of the kind of pattern that the perceived gaze and gesture formed in the communicative context is thus based on the individual analysis and understanding of the indexical relationship that the event may have with respect to the context in which occurred.

6. Conclusion and future work

Natural language communication is intrinsically multimodal, and interlocutors effectively use the different modalities to coordinate and control their actions. For instance, eye-gaze functions as an important signal for the

interlocutors to manage turn-taking and feedback, while gesturing is effectively used for the emphasis and coordination of interaction. This paper has focussed on such signalling processes, and contributed especially to the role of eye-gaze in turn-taking and giving feedback. The method has been a hybrid approach, which unifies signal-level data analysis with human data annotation and enables conversational phenomena to be studied both from the bottom-up view-point of observable data and from the top-down view-point of human perception. It allows comparison of signal-level analysis of the video with the human interpretation of what happens on the video, along the lines described in (Jokinen, 2009b).

Eye-tracking technology has provided new possibilities to study human interaction and to obtain data on the use and function of the interlocutors' gazing in conversational settings. The paper describes how eye-tracker data can be effectively used in the analysis of gaze-paths and be included in multimodal dialogue annotation. The analysis and classification tests contributed to the modelling of interaction in multiparty setting, and although the data is rather small, it was possible to draw some conclusions concerning the differences between eye-gaze in dialogues and in multiparty conversations. For instance, it was noticed that in multiparty conversations, eye-gaze functions as an important cue to indicate the interlocutors' focus of attention, but it does not have such a definite role in turn-taking and feedback as in two-party dialogues. It is hypothesized that the reason for this is the larger shared space in which the interlocutors have to operate: their focus of attention and their responsibility of the interaction is distributed among all the conversational partners, and thus to signal their communicative intentions, they also need more visible signals in order to catch the partners' attention. For this purpose, head turning and hand gesturing could be used. On the basis of the current data, however, only head movement seems to be an indicative signal for turn-taking, whereas hand gesturing was found to be less correlated with the interlocutors' turn taking and turn accepting behaviour. The results were then also related to the nature of non-verbal communication in general. It is assumed that non-verbal elements are indexical signs rather than meaning carrying symbols and their interpretation is linked to the whole context in which they occur. Thus the preference for head turns over hand-gestures as turn-taking signals can be related to the fact that head turns indicate changes in the partner's gazing and thus form a basis for mutual eye-gaze, unlike hand gestures which mainly catch the partner's attention but do not necessarily lead to mutual gaze.

Concerning the future work, it was already mentioned in Section 3 that in order to investigate possible effects of the constraints of the eye-tracker technology on the interlocutors' communicative behaviour it would be useful to compare the interlocutors' gesturing and body movements in the current data with situations where no eye-tracker is used. Moreover, to substantiate hypotheses about possible differences in non-verbal communication

between the Japanese and English speaking groups, it is necessary conduct intercultural comparison studies which provide many interesting research topics. Work in this respect has already started concerning nodding, as well as backwards and forwards body movements.

The data also contains conversations between interlocutor groups where the participants either know each other or do not know each other, and thus it is possible to compare the interlocutors' non-verbal communication depending on their mutual familiarity. It is expected that free conversations with unfamiliar partners are still more formal in style than chatting with friends, and that the difference is also reflected in the speakers' hand gesturing and body movements: these are fewer, less obtrusive, and intrusive than those used in familiar settings.

Yet another future research topic in these lines concerns differences between group discussions and two-party dialogues, especially in situations where the participants are unfamiliar with each other and they have to introduce themselves and get to know each other better. This kind of comparison is planned in relation to the Nordic NOMCO corpus (Paggio et al., 2010). The first-encounter dialogues in the NOMCO corpus have no eye-tracker information, but the annotation schemes for the two corpora are related and provide a systematic basis for comparing the speakers' behaviour along various multimodal aspects.

7. Acknowledgements

The eye-gaze data collection and analysis was done when the author was NICT Visiting Scholar at Doshisha University, Kyoto, Japan. I would like to thank Prof. Yamamoto for providing an excellent environment to work, Prof. Nishida for assistance in many practical issues, and students for help in the data collection and annotation.

8. References

- Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., Paggio, P. (2007). The MUMIN Coding Scheme for the Annotation of Feedback, Turn Management, and Sequencing Phenomena. *Multimodal Corpora for Modelling Human Multimodal Behaviour. Special issue of the International Journal of Language Resources and Evaluation*, 41(3-4), 273-287.
- Argyle, M., Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.
- Bavelas, J. B. (2005). Appreciating face-to-face dialogue. In AVSP-2005, 1.
- Ishii, R., Nakano, Y. (2008). Estimating User's Conversational Engagement Based on Gaze Behaviors. In H. Prendinger, J. Lester, & M. Ishizuka (Eds.): *IVA 2008*, LNCS 5208. Berlin Heidelberg: Springer-Verlag. pp. 200–207.
- Jacob, R.J.K., Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary). In J. Hyona, R. Radach, & H. Deubel, (Eds.): *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam: Elsevier Science. pp. 573-605.
- Jokinen, K. (2009a). *Constructive Dialogue Management. Rational Agents and Speech Interfaces*. Chichester: John Wiley.
- Jokinen, K. (2009b). Gaze and Gesture Activity in Communication. In C. Stephanidis (Ed.): *Universal Access in Human-Computer Interaction*. Proceedings of the 5th International Conference of UAHC, held as Part of HCI International, San Diego, CA.
- Jokinen, K., Nishida, M., Yamamoto, S. (2009). Eye-gaze Experiments for Conversation Monitoring. In: *Proceedings of the IUCS'09 conference*. Tokyo: ACM.
- Jokinen, K., Nishida, M., Yamamoto, S. (2010a). On Eye-gaze and Turn-taking. In *Proceedings of the Workshop "Eye-gaze in Intelligent Human-Machine Interaction", International Conference on Intelligent User Interfaces*. Hong Kong.
- Jokinen, K., Nishida, M., Yamamoto, S. (2010b). Collecting and Annotating Conversational Eye-Gaze Data. In *Proceedings of the LREC Workshop on Multimodal Corpus Collection and Coding*, Malta.
- Kendon, A. (1967). Some Functions of Gaze Direction in Social Interaction. *Acta Psychologica* 26: 22–63.
- Kipp, M. (2001). Anvil – A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of Eurospeech*. pp. 1367–1370.
- Lee, J., Marsella, T., Traum, D., Gratch, J., Lance, B. (2007). The Rickel Gaze Model: A Window on the Mind of a Virtual Human. In *Proceedings of the 7th International Conference on Intelligent Virtual Agents*. Springer Lecture Notes in Artificial Intelligence 4722. Berlin Heidelberg: Springer-Verlag. pp. 296-303.
- Majaranta, P., Räihä, K. (2002). Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*. ETRA '02, New York, NY, pp.15-22.
- Nakano, Y., Nishida, T. (2007). Attentional Behaviours as Nonverbal Communicative Signals in Situated Interactions with Conversational Agents. In Nishida, T. (Ed.) *Engineering Approaches to Conversational Informatics*, John Wiley.
- Paggio, P., Allwood, J., Ahlsén, E., Jokinen, K., Navarretta C. (2010). The NOMCO multimodal Nordic resource – goals and characteristics. In *Proceedings of the LREC conference*, Malta.
- Trevarthen, C. (1984). Emotions in Infancy: Regulators of Contact and Relationships with Persons. In K. Scherer & P. Ekman (eds.) *Approaches to Emotion*. Hillsdale: Erlbaum.
- Vertegaal, R., Weevers, I., Sohn, C., Cheung, C. (2003). GAZE-2: conveying eye contact in group videoconferencing using eye-controlled camera direction. In *Proceedings of CHI 2003*. Fort Lauderdale, TX: ACM Press.

