ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Gaze Patterns in Turn-Taking

Catharine Oertel[1], *Marcin Włodarczak*[2],
Jens Edlund[1], *Petra Wagner*[2], *Joakim Gustafson*[1]

[1]Department of Speech, Music and Hearing, KTH, Stockholm, Sweden
[2] Faculty of Linguistics and Literary Sciences, Bielefeld University, Bielefeld, Germany

catha@kth.se, mwlodarczak@uni-bielefeld.de,
edlund@speech.kth.se, petra.wagner@uni-bielefeld.de, jocke@speech.kth.se

Abstract

This paper investigates gaze patterns in turn-taking. We focus on differences between speaker changes resulting in silences and overlaps. We also investigate gaze patterns around backchannels and around silences not involving speaker changes.
Index Terms: gaze, turn-taking, dialogue, inter-speaker coordination

10.21437/Interspeech.2012-132

1. Introduction

Gaze is one of the strongest and most studied visual cues in face-to-face interaction, and has been associated with a variety of functions, such as managing attention in dialogue partners [1], expressing intimacy and exercising social control [2], highlighting the information structure of the propositional content of speech [3] and coordinating turn-taking [4, 5]. In this study we are concerned with the last one. The fundamental gaze patterns related to turn negotiation were discussed in Kendon's seminal 1967 paper [5]. Here, we extend this work and that of Kendon's successors in two ways. Firstly, given the pervasiveness of overlaps in spontaneous conversations, e.g., [6], we compare gaze behaviour for speaker transitions in overlap, in silence, and in no perceptual overlap or silence [7]. We also include an analysis of gaze patterns in the vicinity of backchannels. Secondly, Kendon's data, while highly informative, has the disadvantage of having been recorded at a low frame rate of two frames per second. We provide a temporally more fine-grained account of the *dynamics of change* of these gaze patterns.

2. Background

The fundamental gaze patterns related to turn negotiation were discussed in Kendon [5], who demonstrated that speakers look away at turn beginnings and look back at their partners towards turn endings. Kendon also identified listeners' gaze at the speaker as an attention signal and looking away as an agreement signal. Tentative evidence that "floor fights" are characterised by an increase in mutual gaze was also found. Bavelas et al. [8] found that gaze patterns used to coordinate collaborative responses in dialogue are often preceded by the speaker gazing at the listener, resulting in short periods of mutual gaze (in their paper referred to as *gaze windows*) broken by the listener looking away shortly afterwards. A similar approach to studying gaze in interaction was adopted by Cummins [9]. He found that many of the gaze patterns vary substantially from one speaker pair to another and should be considered "a dynamic feature of a specific conversational situation". His results are in line with earlier studies on gaze coordination in dialogue which demonstrated its dependence on factors such as the established common ground and mutual knowledge [10].

In recent years such findings have been applied to human-machine interaction in order to make avatars appear more human-like, e.g. [11, 12].

3. Corpus

A subset of the IFADV corpus [13] was used. The IFADV corpus is a video corpus of spontaneous Dutch dialogues. All participants knew each other prior to the recordings; they were either good friends or have worked together for a long time. The participants were seated at opposite ends of a table, facing each other. Two cameras were used, each capturing the face and upper body parts of a participant at a frame rate of 25 frames/s. For this study we used a subset of seven dialogues, approximately 15 minutes each, with a total duration of 105 minutes.

4. Data annotation and analysis

Gaze was annotated according to the scheme proposed by Cummins [9]. The data was annotated manually on a frame-by-frame basis. For each frame a binary distinction was made between looking at the face of the interlocutor (g) and looking away (x). Gaze of both interlocutors was annotated. Additionally, turn boundaries as well as backchannels were marked. Turn-internal silences were considered to be part of a turn and not in-

The first two authors contributed to the paper equally.

cluded in the annotation. Consequently, silences bounded by speech from the same speaker are only present if the speaker released the turn but the other person failed to take it up.

The gaze and utterance annotations were used to identify silences and stretches of overlapping speech with and without speaker change. An additional distinction was made between intervals lasting longer or shorter than 130 ms, which is the reported detection threshold for silences and overlaps in conversation [7]. This resulted in the following inventory (the numbers in brackets indicate counts in a category):

(a) Overlap with speaker change (OV with SC): overlapping speech of at least 130 ms with the incoming speaker continuing after the overlap is resolved (272).

(b) Overlap without speaker change (OV without SC): overlapping speech of at least 130 ms with the original speaker continuing after the overlap is resolved (124).

(c) Silence with speaker change (SIL with SC): silence of at least 130 ms terminated by speech from the incoming speaker (374).

(d) Silence without speaker change (SIL without SC): silence of at least 130 ms terminated by speech from the previous speakers (102)

(e) Overlap with backchannel (OV with BACK): overlapping speech in which the incoming speaker produces a backchannel (817).

(f) Silence with backchannel (SIL with BACK): silence terminated by a backchannel from the incoming speaker (101).

(g) No-gap no-overlap (No-GAP No-OV): overlapping speech or silence with a duration of less than 130 ms (187).

Gaze calculation was carried out in three steps. First, overlap onsets and silence offsets were selected. These points correspond to onsets of incoming speaker's turn (for overlaps and silences with speaker change) or onsets of the previous speaker's continuations (for silences without speaker change). Next, for each participant a binary variable indicating gaze directed at the partner was calculated in 10 ms[1] intervals for three seconds preceding and three seconds following the selected time points. Finally, the proportion of times that a participant was looking at his or her partner at a given time point was calculated for each of the seven categories. The same was carried out for mutual gaze. The procedure is illustrated in Figure 1 and the results are plotted in Figure 2.

Two tailed randomisation tests were used to compare the categories with instances assigned randomly between them and a parameter of interest, e.g. slope, computed in each of 10,000 iterations. *p-values* were calculated as the proportion of values at least as extreme as the observed value.

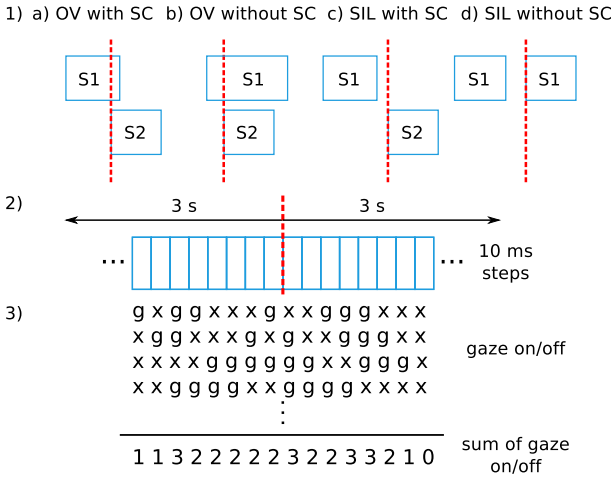

Figure 1: Method for computing gaze patterns. Each row in step 3) corresponds to one instance of SIL/OV for one speaker (e.g. in the case of OV with SC the number of rows = 272)

5. Results

Overlaps with speaker change (Figure 2a) and silences with speaker change (Figure 2c) display a similar pattern characterised by previous speakers looking at their partners and incoming speakers averting their gaze. However, the time course is somewhat different in each category. In overlaps a sharp increase in partner-oriented gaze occurs shortly after the overlap onset and this level is sustained until the overlap resolution. By contrast, for silences the slope is much more gradual and the curve peaks about 1 second after the onset of the incoming speaker's turn. The difference in slopes between the categories in the interval between 0 and 1 seconds is indeed significant at $p < 0.05$.

Incoming speakers' behaviour follows a similar trajectory in both categories. Incoming speakers start to avert their gaze approximately 1 second before the start of the interval in question. The decrease in partner-oriented gaze reaches its minimum shortly after the onset of the incoming speaker's turn. This is also reflected in the gradual decrease in mutual gaze. However, the fall is much greater for non-overlapped speaker changes, which indi-

[1]While the chosen time step is much smaller than the frame rate of the videos, it should be noted that a transition between frames can occur at any time relative to turn boundaries. Consequently, a small time step allowed a more precise identification of gaze switches preceding or following turn boundaries by a value other than a multiple of the frame size.

cates a higher proportion of incoming speakers looking away ($p < 0.05$).

As can be seen in Figure 2b the gaze pattern of overlaps without speaker change is less clear than of overlaps involving a speaker change, possibly due to fewer data points in this category. What should be noted, however, is that, similar to overlaps resulting in a speaker change, there is an increase in previous speakers' partner-oriented gaze directly following the overlap onset (possibly corresponding to the point when the interlocutors notice the overlap). Unlike in overlaps with speaker change, the peak is found before the overlap resolution, the difference in peak location between the categories is not significant ($p > 0.5$).

In silences without speaker change (Figure 2d) previous speakers can be observed to start looking away as early as two seconds before the silence onset. Since in our data this category represents cases when the previous speaker released the turn without the partner taking the floor (see Section 4) the expected pattern for the previous speaker would be similar to that found in silences with speaker change (Figure 2c), in which the previous speaker also yields the turn. However, the last displays quite a different pattern with the previous speaker continuing to gaze at his interlocutor throughout the duration of the silence (significant difference in slope in the interval between -1 and 0 seconds, $p < 0.01$). This might suggest that negotiation of who continues after the gap occurs while the previous speaker is still holding the floor.

Overlaps with backchannels (Figure 2e) and silences with backchannel (Figure 2f) are characterised by a substantial increase in previous speakers' partner-oriented gaze, not observed for any other category. The difference between the minimum and maximum value in the interval between -2 and 2 seconds is significantly greater in (e) compared to (a) ($p < 0.01$), and in (f) compared to (c) ($p < 0.001$). This results in a similar increase of mutual gaze. Additionally, the incoming speaker tends to look away much more when a backchannel is produced in a non-overlapped position ($p < 0.001$). Not surprisingly, the no-gap no-overlap category (Figure 2g) displays the familiar pattern of silences with speaker change characterised by an increase in partner-oriented gaze in the previous speaker and an analogous decrease in the incoming speaker. However, the decrease in incoming speakers' partner-oriented gaze is much smaller ($p < 0.01$) and is not followed by the gradual rise observed in (c).

6. Discussion

The results outlined in the previous section are broadly compatible with Kendon's and Bavelas et al.'s findings. Firstly, speakers were indeed observed to look towards their partners as they are about to release their turn and look away at the start of a new turn. Secondly, backchannels are indeed associated with an increase in mutual

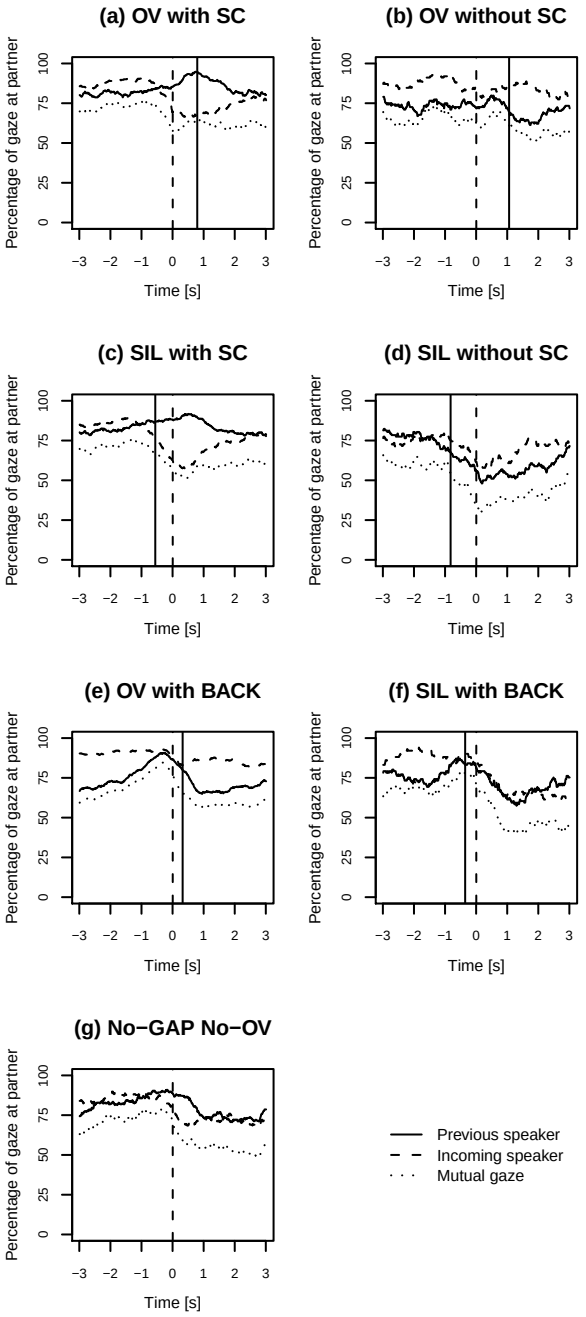

Figure 2: Proportions of partner-oriented gaze and mutual gaze for (a) overlaps with speaker change, (b) overlaps without speaker change, (c) silences with speaker change, (d) silences without speaker change, (e) overlaps with backchannel, (f) silences with backchannels, (g) no-gap no-overlap calculated in 10 ms time steps for 3 second intervals preceding and following overlap onsets and silence offsets (marked with the vertical dashed line). The vertical solid lines represent mean silence and overlap durations.

gaze directly preceding the onset of the feedback expression. We also find a greater increase in the previous speaker's proportion of gaze in backchannels than in speaker changes (with or without overlaps). This could be explained by the fact that listener responses often are visual responses (an eyebrow raise, a head nod or a smile) with or without an accompanying verbal backchannel token. In order to detect these visual cues the speaker has to look at the listener. And since listener responses typically are very short it also leads to the very short peak and sharp decrease of mutual gaze found in Figure 2 (e) and (f). However, neither of these accounts explains the pattern observed for silences without speaker change in our data insofar as they represent cases when the original speaker did release the floor and the other person failed to take their turn. We take the high proportion of turn-holders looking away prior to yielding the floor as evidence of turn negotiation with the speaker modifying his gaze behaviour if his partner does not seem willing to become the next speaker.

In addition, some other patterns not reported in the literature were found. Firstly, overlaps with speaker change are characterised by a sudden increase in previous speaker's partner-oriented gazing following the onset of the incoming speaker's turn. However, given that incoming speakers still tend to look away (albeit to a lesser extent than in silences involving speaker change) it is hard to see how this pattern could correspond to what Kendon described as participants staring "fully at one another" [5].

Secondly, while incoming speakers were observed to look away in all the analysed categories, the extent to which this is the case seems to be somewhat greater for speaker changes than turn continuations, and for in-overlap than out-of-overlap configurations.

Thirdly, silences and overlaps shorter than 130 ms were observed to be similar to the silences with speaker change. This is an expected pattern given that those should be perceived as "smooth" speaker changes. However, unlike in speaker changes accompanied by a perceptible silence, incoming speakers do not tend to look back at their partners after taking the turn.

Lastly, it should be noted that the observed changes in gaze patters extend well beyond the boundaries of the intervals in question. This might be an important finding for the design of conversational agents which could use those as cues for improving the responsiveness and naturalness of their turn-taking.

7. Conclusion

In this study we demonstrated that there are other distinctive gaze patterns in addition to those associated with smooth speaker changes. Speakers' gazing was found to vary according to whether a turn negotiation results in a speaker change or a continuation of turn and whether it coincides with an overlap or with a gap. We also described distinctive gaze pattern for backchannels.

In a future study, the fact that backchannels are cued as far as 2 seconds prior to the actual occurrence of the backchannel might be used for audiovisual synthesis. Additionally, a more fine-grained annotation of turn-internal silences might allow for a detailed analysis of gaze patterns associated with pausing.

8. Acknowledgements

We would like to thank Fred Cummins for granting us access to his annotations. Catharine Oertel is supported by the Swedish Research Council (VR) project Introducing interactional phenomena in speech synthesis (2009-4291). Marcin Włodarczak is supported by the German BMBF-funded "Professorinnenprogramm" FKZ 01FP09105A. The work was funded in part by the Swedish Research Council project: The rhythm of conversation.

9. References

[1] R. Vertegaal, R. Slagter, G. van Der Veer, and A. Nijholt, "Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes," in *SIGCHI Conference on Human Factors in Computing Systems*, 2001, p. 308.

[2] C. Kleinke, "Gaze and eye contact: A research review," *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100, 1986.

[3] J. Cassell, *Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents.* Cambridge, MA: MIT Press, 2000, pp. 1–27.

[4] S. Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personality and Social Psychology*, vol. 23, pp. 283–292, 1972.

[5] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22–63, 1967.

[6] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 30, no. 4, pp. 555–568, 2010.

[7] M. Heldner, "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," *Journal of Acoustical Society of America*, vol. 130, no. 1, pp. 508–513, 2011.

[8] J. B. Bavelas, L. Coates, and T. Johnson, "Listener responses as a collaborative process: The role of gaze," *Journal of Communication*, vol. 52, no. 3, pp. 566–580, 2002.

[9] F. Cummins, "Gaze and Blinking in Dyadic Conversation: A study in Coordinated Behavior Among Individuaks," *Language and Cognitive Processes*, in press.

[10] K. Shockley, D. C. Richardson, and R. Dale, "Conversation and coordinative structures," *Topics in Cognitive Science*, vol. 1, no. 2, pp. 305–319, 2009.

[11] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. And Stone, "Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents," in *ACM SIGGRAPH 94*, 1994, pp. 413–420.

[12] S. Al Moubayed, J. Edlund, and J. Beskow, "Taming Mona Lisa: communicating gaze faithfully in 2D and 3D facial projections," *ACM Transactions on Interactive Intelligent Systems*, vol. 1, no. 2, p. 25, 2012.

[13] R. van Son, W. Wesseling, E. Sanders, and H. van Den Heuvel, "The IFADV corpus: A free dialog corpus," in *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, Marrakech, Morocco, 2008, pp. 501–508.