When Robot Gaze Helps Human Listeners: Attentional versus Intentional Account

Maria Staudte & Matthew W. Crocker
Department of Computational Linguistics
Saarland University
Saarbrücken, Germany
{masta, crocker}@coli.uni-saarland.de

Abstract

Previous research has shown that listeners exploit speaker gaze to objects in a shared scene to ground referring expressions, not only during human-human interaction, but also in human-robot interaction. This paper examines whether the benefits of such referential gaze cues are best explained by an attentional account, where gaze simply serves to direct the listeners visual attention to an object immediately prior to mention, or an intentional account, where speaker gaze is rather interpreted as revealing the referential intentions of the speaker. Two eye-tracking studies within a human-robot interaction setting are presented which suggest that close temporal synchronization of speaker gaze and utterance is not necessary to facilitate comprehension, while the order of gaze cues with respect to order of mentioned references is. We interpret this as evidence in favor of an intentional account.

Keywords: human-robot interaction; gaze; visual attention; referring expressions

Introduction

Gaze has been widely studied as an indicator for overt visual attention during language processing. Previous studies revealed that speakers look at entities shortly before mentioning them (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998), while listeners rapidly inspect objects as they are mentioned (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This shows that gaze during situated language production and comprehension is tightly coupled with the unfolding speech stream, reflecting both speakers' intentions and listeners' understanding on-line. In face-to-face communication, the speaker's gaze to objects in a shared scene provides the listener with a visual cue to the speaker's focus of (visual) attention (Emery, 2000; Flom, Lee, & Muir, 2007). By revealing a speaker's focus of visual attention, such gaze cues potentially offer the listener valuable information to ground and disambiguate referring expressions, to hypothesize about the speaker's communicative intentions and goals and, thus, to facilitate comprehension (Hanna & Brennan, 2007).

In human-robot interaction, robot gaze that was synchronized with speech in a human-like manner has been shown to be similarly useful for grounding and resolving spoken references (Staudte & Crocker, 2009b). Further evidence supported the hypothesis that the utility of robot gaze originates from people's inferences of referential intentions from gaze (Staudte & Crocker, 2009a). However, it remained an open question whether such human-like synchronization of gaze and spoken references is necessary for gaze to be beneficial.

Firstly, we hypothesize that people indeed infer referential intentions from robot gaze cues. And secondly, we hypothesize that this assignment of *intentional* states makes the utility of gaze relatively flexible with respect to temporal synchronization. That is, despite a substantial shift of gaze cues with respect to corresponding speech cues, the conveyed intentions of the speaker may still facilitate utterance comprehension. If, in contrast, gaze is only a purely visual cue that happens to directs listeners' *visual attention* to an object which is then mentioned, we hypothesize that close temporal synchronization would be necessary for any benefit of gaze. We present evidence from two experiments supporting an intentional account of processing and interpreting robot gaze.

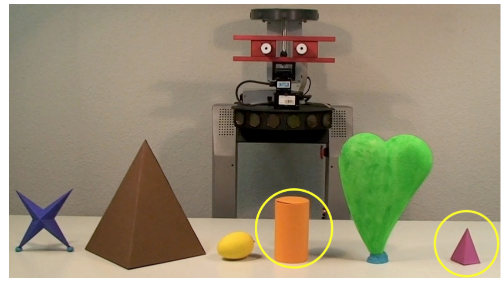
Does robot gaze reflect referential intentions?

In previous experiments, Staudte and Crocker (2009b) showed that people follow and use robot gaze, similar to human gaze, and faster resolve referring expressions in the robot's utterance only when the gaze cue identified the actually mentioned object. Two different explanations of this result are conceivable. The influence of robot gaze could possibly be explained in terms of a purely "bottom-up" process: Robot gaze draws attention to one object (cf. Langton et al., 2000, for reflexive orienting in response to gaze cues) and the utterance subsequently draws attention to the same (congruent) or another object (incongruent). Thus, incongruent gaze elicits an *additional* shift of visual attention before utterance comprehension is completed. This additional shift could simply add to the total time needed to comprehend and respond, thus, accounting for an increase in response times (we will refer to this as the *Visual Account*). However, the effect of robot gaze could also be explained in terms of a (mis)match in expectations (elicited by robot gaze) and the actual utterance. Previous studies on the interpretation of human gaze have revealed that gaze is an extremely versatile cue which reflects attentional states as well as mental states such as goals, desires and intentions (Baron-Cohen, 1995). We therefore hypothesized that people's use of robot gaze may also be driven "top-down", by the belief that robot gaze also reflects attentional and intentional states and, thus, reveals what the robot intends to mention (*Intentional Account*). That is, participants may have thought that the robot attended to one object because it intended to mention it and, therefore, an incongruent reference would have led to a revision in referential expectations which slowed people. In a follow-up experiment, participants were asked to correct false robot utterances and were free to decide which objects they mentioned in their correction sentence (Staudte & Crocker, 2009a). Participants' responses suggested that their understanding about what the robot had originally intended to say was indeed affected by its gaze.

Another way to potentially distinguish between Intentional and Visual Account is to consider the relevance of temporal synchronization of gaze and speech cues. Recent findings from a study on the influence of *indirect*, human speaker gaze on utterance comprehension suggest that there is only limited flexibility in the requirement of synchronization of such a gaze cue with speech, while maintaining its utility for the listener (Kreysa, 2009). Kreysa (2009) found that gaze cues with a small shift with respect to their natural temporal co-occurrence still facilitated task completion whereas a greater shift (by more than 2 sec) was no more beneficial than random cues. Interestingly, the importance of synchronization between gaze and speech may illuminate the nature of gaze influence. On the Visual Account, the influence of gaze is attributed to the induced attention shift towards the right object at the right time such that changes in the temporal synchronization of gaze and speech should clearly affect the utility of gaze. Under the Intentional Account, in contrast, people would interpret robot gaze with respect to the robot's intentional states such that synchronization would not be critical. Understanding someone's (referential) intentions should persist and influence utterance comprehension as long as they seem relevant.

While Kreysa's results (2009) suggest that the effect of human gaze cues on utterance comprehension is flexible to some extent, gaze cues used in her studies were indirect and not necessarily qualitatively equal with the direct perception of speaker gaze. Depending on how people perceive speaker gaze compared to Kreysa's cursor, two different behaviors in response to substantially shifted robot gaze is possible: Robot gaze may be similar to a gaze cursor, a *visual cue* that may (reflexively) direct attention and, thus, is only helpful for processing referring expressions when it occurs within a short time window around the spoken reference. A substantial shift of gaze relative to speech would result in longer response times than the original synchronization. Alternatively, speakers' looks towards an object may be perceived as more *intentional* than a gaze cursor and as more robustly assigning relevance to the object in focus (similar to human gaze). Participants may persistently maintain and use this information when it seems relevant, leading to equal response time for shifted and synchronized gaze. Equally, non-congruent gaze cues may thus – even when shifted to precede the utterance – disrupt comprehension and cause slower response times.

We present results from two experiments which suggest that the utility of gaze is *not only* a matter of attention cueing to the right object at the right time. Rather people seem to interpret robot gaze as an indicator to the robot's referential intentions, leading to a persistent influence of gaze.



Original
Sync: <c>"The cylinder is taller than the <p> pink pyramid."
Prec: <c><p>"The cylinder is taller than the pink pyramid."

Reverse
Sync: <c>"The pink pyramid is shorter than <p> the cylinder."
Prec: <c><p>"The pink pyramid is shorter than the cylinder."

Figure 1: Sample scene from experiments, with original/reversed sentences and synchronized/preceding robot gaze (first at cylinder (<c>), then at pyramid (<p>)).

Experiment 1

In this study, we investigated whether referential robot gaze needs to be temporally synchronized with speech (in the way human gaze is synchronized) in order to be beneficial, or whether robot gaze conveys referential intentions that have a more persistent effect on utterance comprehension. Thus, we manipulated synchronization in two ways. While robot gaze was always directed to the mentioned objects, we manipulated the factor Order of Mention (sequence of mentioned objects crossed with sequence of 'gazed at' objects) which led to original (coherent) or reverse order of references (see Figure 1). The second factor, Synchronization, manipulated the temporal delay between gaze/visual references and corresponding linguistic references.

Method

Participants Thirty-two native speakers of German, mainly students enrolled at Saarland University, took part in this study (26 females). All participants reported normal or corrected-to-normal vision.

Materials We created 1920x1080 resolution video-clips showing a PeopleBot robot (kindly provided by the CogX-project, http://cogx.eu) onto which a pan-tilt unit was mounted. This pan-tilt unit carried a stereo camera which appeared as the head and eyes of the robot. The video-clips each showed a sequence of camera-movements consecutively towards the central object and then the peripherally located object. The utterance was a synthesized German sentence using the Mary TTS system (Schroeder & Trouvain, 2001).

In these videos we manipulated two factors: Order of Mention (original, reverse) and Synchronization (synchronized, preceding), so each item appeared in four conditions. The temporal delay between gaze and speech was roughly 5.3 seconds in the preceding condition and 1 second in the synchronized condition. A sample scene is given in Figure 1 as well as examples for each type of sentence order and robot gaze synchronization. In condition original-synchronized, gaze and speech cues were coherent and synchronized in a human-like manner while the condition reverse-synchronized showed cues that were concurrent but reverse to each other.

Eight lists of stimuli were created, accounting for four experimental conditions and their counter-balanced versions. In addition to 24 items, 36 fillers were shown such that participants saw a total of 60 trials. The order of item trials was randomized for each participant individually and items were always separated by at least one filler.

Procedure An EyeLink II head-mounted eye-tracker monitored participants' eye movements on a 24-inch monitor. Before the experiment, participants received written instructions about the experiment procedure and task: They were asked to attend to the presented videos and judge whether or not the robot's statement in each was valid with respect to the scene. In order to provide a cover story for this task, participants were further told that the results were used as feedback in a machine learning procedure for the robot.

Analysis Videos were segmented into Interest Areas (IAs). That is, each video contained a region labeled "NP2 referent" which marked the object mentioned last in the robot utterance (i.e., before sentence validation was possible). Further, we recorded and analyzed participant fixations on this area. The speech stream was segmented into two Interest Periods (IPs). IP1 was defined as the 1000ms period ending at the onset of the second noun (in NP2). Importantly, it contained the robot's gaze towards the target object as well as verbal content preceding the target noun. IP2 was defined as the 700ms period beginning with noun onset in NP2. These IPs roughly segmented the sentences as follows: *"The cylinder is taller [than the pink]$_{IP1}$ [pyramid]$_{IP2}$"*. Defining IP1 and IP2 in this way made it possible to distinguish once again between gaze-mediated inspections in IP1 and utterance-mediated inspections in IP2. Trials that contained at least one *beginning* inspection towards an IA within an IP (coded as "1") were contrasted with trials that did not contain an inspection in the same slot ("0"). As a result, mean values represent inspection probabilities for a given IA and IP. For inferential analyses, we considered inspections on the NP2 referent as well as response time, recorded from NP2-noun onset to the moment of the button press. The analyses were carried out using mixed-effect models from the lme4 package in R and Chi-Square tests to asses the contribution of a predictor through model reduction (Baayen, Davidson, & Bates, 2008; Bates, 2005).

Results

Eye movements Mean inspection probabilities for the NP2 referent are depicted in Figure 2. Note, that the manipulation of Order of Mention coincided with a difference in location of the NP2 referent. That is, in original order, the NP2 referent is located in the periphery of the table (pink pyramid), while in reverse order it is located in the center of the scene (cylinder), as depicted in Figure 1. Results from inferential statistics on inspection data are reported in-text where necessary and are otherwise omitted due to space limitations.

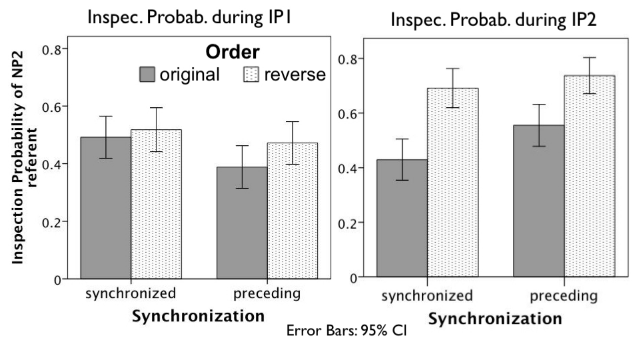


Figure 2: Inspection probability of NP2 referent in Exp1, for all conditions in IP1 (left graph) and IP2 (right graph).

In IP1, model reduction revealed a main effect of Synchronization on inspections on the NP2 referent ($\chi^2(1) = 4.03, p < .05$). People inspected the NP2 referent with lower probability when robot gaze preceded the utterance. Moreover, we did not observe a main effect for Order of Mention, i.e., people inspected the NP2 referent equally often irrespective of where this referent was located (centrally, peripherally) or whether the robot concurrently fixated this object. Interestingly, people were equally likely to inspect the NP2 referent in condition original-synchronized (when robot gaze identifies the actual NP2 referent) as in condition reverse-synchronized (when gaze does not). This result indicates that people may use the already mentioned reference (NP1) and the available visual cues to, at least *visually*, anticipate the NP2 referent even when cues were reversed.

In IP2, we observed a somewhat different inspection pattern. Order of Mention had a main effect on inspection probability ($\chi^2(1) = 35.67, p < .001$) such that participants inspected the (mentioned) NP2 referent significantly more often in the reverse condition than in the original (coherent) order condition. That is, when the robot fixated the peripheral object during IP1 and then mentioned the other, central object in IP2 (<central cylinder>"The pink pyramid is taller than <periph. pyramid> the cylinder.") participants were more likely to inspect the object mentioned in NP2 (centrally located cylinder) than in original order.

There are two possible explanations for these high probabilities of inspecting the NP2 referent in reverse order: Either participants inspected this central object more often because it was more salient due to its central location, predicting easy and quick reference resolution. Alternatively, the increased inspections on the NP2 referent in reverse condition reflect difficulty to resolve the reference as it includes conflicting information (gaze identified the pyramid while the mentioned noun referred to the cylinder). Thus, the response time results should reveal which of the two explanations is more likely.

Response Time Model reduction showed that Synchronization had no effect on response times. That is, participants were equally fast to determine the validity of the robot statement in synchronized and preceding conditions. Since no interaction between the two factors Synchronization and Order of Mention was observed, we excluded Synchronization as a predictor from our linear mixed-effects model. Model reduction further revealed a main effect of Order of Mention ($\chi^2(1) = 45.19, p < .001$, see also Figure 3 for averages).

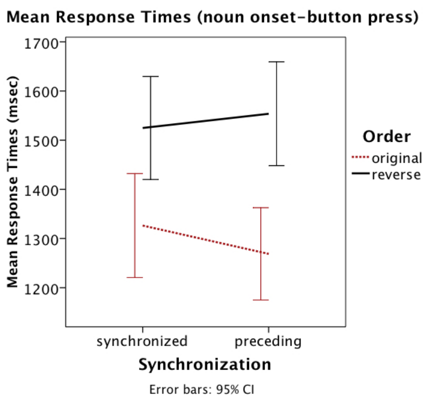


Figure 3: Avg. response times in all four conditions (Exp1).

The finding that Synchronization did not affect response time while Order of Mention did, cannot be explained by the Visual Account since both manipulations made robot gaze direct people's visual attention to relevant objects at non-synchronized points in time – and always prior to the last referring expression (NP2). Instead, the Intentional Account seems to provide more appropriate explanations for these results: The precise temporal synchronization is not crucial for people to interpret and use robot gaze as a cue to the robot's intentions. The inferred (referential) intentions, however, are expected to be executed in the *same order* as they were indicated by robot gaze. Thus, reversed order, even in the case of preceding gaze, slows people in utterance comprehension.

Discussion

By manipulating the order of references in the sentence, the location of the NP2 referent was effectively also manipulated. Since the center of the scene is the most salient area, this may have affected the effort needed to resolve a referring expression which identified the central object compared to one that identified an object in the periphery of the scene. Since in the reverse order condition, NP2 identifies the central object, this appears to benefit reference resolution given that the central object is most salient. However, response time results revealed that people were in fact *slower* in reverse order to judge sentence validity, compared to original order. This effect of Order of Mention suggests that reverse order was indeed more difficult to process than originally ordered cues, supporting the interpretation that people's increased inspections on the NP2 referent reflected *increased effort* to resolve the reference (due to conflicting information).

Experiment 2

In this experiment we further investigated whether the order of referring expressions in the robot utterance (accompanied only by neutral gaze) affects how fast people resolve these expressions and validate the utterance. Original and reverse sentence order were, thus, paired with neutral robot gaze and compared. This baseline condition showing neutral gaze allowed us not only to determine any effects of sentence order itself, but also to assess the actual benefit of original order versus a potentially disruptive effect of reversed gaze and speech cues. Since the temporal shift between synchronized and preceding condition did not affect people's responses, we did not include a preceding gaze condition again. We manipulated Order of Mention (*original*, *reverse*) and Synchronization (*synchronized*, *neutral*).

Synchronized robot gaze was again always directed first to the central object and then to the peripheral object. Using the sample sentence from the previous experiment "The orange cylinder is taller than the pink pyramid", the robot would first look at the cylinder and then to the peripherally located pyramid. The neutral gaze condition showed an initial glance down at the scene before the robot looked straight ahead and began to speak. We included an additional adjective for the central object (the "*orange* cylinder") in order to make sentences completely symmetric in both sentence orders. This symmetry also allowed us to change the onset for response time recordings from NP2-noun onset to NP2-adjective onset. Since the adjective already uniquely identifies the referent this most appropriately captures actual response time. Otherwise sentences and scenes were similar to the material used in Experiment 1.

Method

Participants & Procedure Thirty-two native speakers of German and mostly students at Saarland University took part in this study (21 females). All reported normal or corrected-to-normal vision. Task and Procedure were identical to Experiment 1.

Materials The manipulation of Order of Mention (original, reverse) and Synchronization (synchronized, neutral) resulted in four conditions. A set of 20 items was used as well as a set of 32 fillers which were evenly distributed across conditions. Participants therefore saw a total of 52 trials.

Analysis IAs used in this experiment were identical to those in Experiment 1. IP1 was again defined to begin 1,000ms prior to noun onset (in NP2). However, in this experiment IP1 did not stretch to noun onset but already ended with *adjective onset*. Thus, IP1 had no fixed duration but an average length of 600ms. This shortening of IP1 was done to incorporate the fact that the prenominal adjective already uniquely identified the referent. Consequently, IP2 was defined to stretch from adjective onset to 700ms after noun onset and had a mean duration of 1,100ms. Thus, sentences were segmented as follows: *"The cylinder is taller [than the]*$_{IP1}$

[pink pyramid]$_{IP2}$". Defining IP1 and IP2 in this way made it possible to distinguish once again between gaze-mediated inspections in IP1 (before the linguistic reference in NP2) and utterance-mediated inspections in IP2 (taking into account that the color adjective linguistically identifies the NP2 referent). Moreover, response time was defined to begin with NP2-*adjective onset* instead of the previously used noun onset for the same reason, that is, accounting for the nominal adjective as already identifying the final referent.

Results

Eye movements Mean probabilities for inspecting the NP2 referent are given in Figure 4. In IP1, both Order of Mention and Synchronization had main effects on inspection behavior (Synchronization: $\chi^2(1) = 5.83, p < .05$ and Order of Mention: $\chi^2(1) = 24.90, p < .001$). Participants generally inspected the NP2 referent more frequently when gaze was synchronized than when it was neutral. Moreover, model reduction revealed a significant interaction of the two predictors Order of Mention and Synchronization ($\chi^2(1) = 14.08, p < .001$). That is, the effect of Order of Mention varied depending on the Synchronization: Firstly, the neutral gaze condition reveals that Order of Mention by itself affected people's visual attention. In the reverse-neutral condition the NP2 referent was inspected significantly more often than in original-neutral. We argue that this effect is due to the NP2 referent being central and being additionally highlighted as the robot initially looked downwards. Secondly, the graph also reveals that the peripherally located object (NP2 referent in original order) was inspected more often when gaze was synchronized (original-synchronized) than when it was neutral (original-neutral), suggesting that a gaze cue in original (coherent) order helped people to visually anticipate the NP2 referent. In contrast, gaze cues in reverse order did not affect the inspections on the NP2 referent (central object) compared to reverse-neutral. Instead, the NP2 referent was rather frequently inspected in reverse order even when robot gaze was neutral throughout the utterance. This indicates that the central object was indeed more salient than the peripheral object.

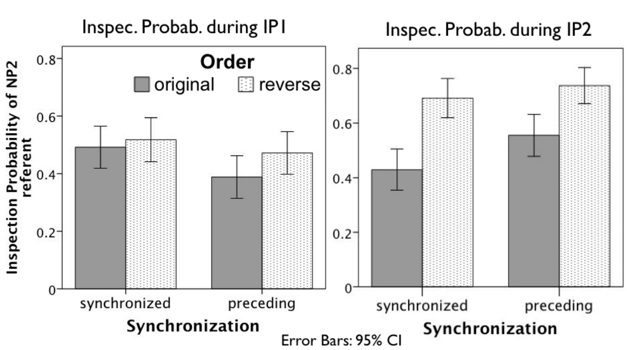


Figure 4: Inspection probability on NP2 referent in Exp2, for all conditions in IP1 (left graph) and IP2 (right graph).

In IP2, Order of Mention had a main effect on inspection probabilities ($\chi^2(1) = 51.99, p < .001$). That is, during NP2 noun mentioning, people inspected the NP2 referent more frequently in reverse order than in original order. As in Experiment 1, this suggests that people visually attended more closely to the mentioned object when the referring expression required more effort to be resolved.

Response Time Model reduction revealed a significant interaction of both predictors, Order of Mention and Synchronization ($\chi^2(1) = 16.85, p < .001$). Consequently, both predictors were included in the model fitted to response times. This model is specified by our dependent variable response time, the two predictors Order and Synchronization, and two random factors accounting for subject and item variation ($DV \sim Predictor1 \times Predictor2 + randomFactors$, see also Table 1). Both factors had a marginal main effect, however, the interaction is clearly more relevant for interpretation as is explained below. Firstly, pairwise comparisons reveal the following significant differences: Between reverse-neutral and reverse-synchronized ($p < .001$), reverse-synchronized and original-synchronized ($p < .05$), reverse-neutral and original-neutral ($p < .001$) and a marginally significant difference between original-synchronized and original-neutral ($p = .07$). These results suggest that order of references in a sentence indeed affected participant behavior, as already suggested by the inspection data in IP1. The response times in both neutral conditions show that people were significantly faster to validate the robot's utterance in the reverse-neutral condition than in original-neutral. This result is consistent with the findings of visual anticipation of the NP2 referent (for neutral gaze), i.e., when order was reversed people anticipated the NP2 referent, when order was original they hardly did. This suggests that reverse order of mention was generally *easier* to process than original order of mention. However, synchronization of gaze cues reversed this effect: Participants were significantly slower when gaze was synchronized and in reverse order (resulting in concurrent but conflicting referential cues) than when gaze was synchronized and in original order (concurrent and coherent order of cues).[1]

[1] The response time pattern in Experiment 2 was largely independent of the chosen onset. That is, results were qualitatively equal for starting recording at NP2-adjective onset or at NP2-noun onset.

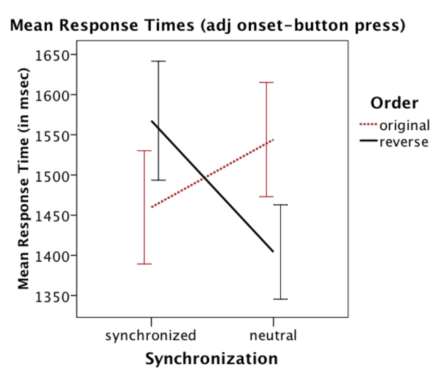


Figure 5: Avg. response times in all four conditions (Exp2).

Table 1: Model fitted to response time data. The last column shows p-Values calculated through Monte-Carlo-sampling.

Predictor	Coeff.	*SE*	t-value	pMCMC
(Intercept)	1475.79	55.19	26.741	<.001
Order-reverse	96.24	40.36	2.384	<.05
Synchr.-neutral	68.89	39.60	1.740	.075
reverse.:neutral	-230.67	55.94	-4.124	<.001

$Model : RT \sim OrderofMention \times Synchronization + (1|subject) + (1|item)$

Discussion and Conclusion

Results from Experiment 1 suggest that gaze cues are equally beneficial when preceding the spoken references as when they occur concurrently. However, the interpretation with regard to the influence of cue ordering was difficult as the manipulation of order was potentially confounded with the sentence order (i.e., referent location in the scene). This was addressed by adding a neutral gaze condition in Experiment 2 which revealed that reverse sentence order was *easier* to process than original sentence order. Thus, despite the advantage of reverse sentence order, synchronizing (reverse) robot gaze cues disrupted people whereas adding original (and coherently) ordered gaze cues to original sentence order significantly enhanced response time of this sentence order. The results for synchronized robot gaze may therefore be interpreted with respect to gaze and speech cue synchronization only: Synchronizing (reversed) gaze cue with reverse order of mention *increased* response times, while synchronizing (coherent) gaze cues with original order of mention *reduced* response times, when each is compared to its neutral gaze baseline.

The presented results thus suggest that large temporal shifts of robot gaze with respect to its 'natural' synchronization do not substantially affect the utility of the gaze cues whereas the order of the cues does. This contradicts the predictions derived from the Visual Account. The Intentional Account, in contrast, provides a plausible explanation for these results: The precise temporal synchronization is not critical since people interpret and use robot gaze as a cue to the robot's intentions. This may also explain why Kreysa (2009) found that substantial temporal shifts reduce the gaze cursor's utility while robot gaze and speech synchronization, in contrast, appears rather flexible. The order of cues, however, affects the the utility of robot gaze since the order of inspections reflects the speaker's intentions regarding order of mention. Thus, people seem to expect that the inferred referential intentions be realized in the corresponding order (Griffin & Bock, 2000). If this expectation is not met, gaze cues may even disrupt comprehension, as the comparison with neutral gaze suggests. Consequently, the presented evidence for a flexible use of robot gaze during utterance comprehension further supports the hypothesis that people assign attentional and intentional states to the robot. Thus, future research could use such a robot interaction setting to generally address questions such as the extent to which people infer intentions or other information from their partner's gaze. While it is not entirely clear whether robots and agents provide an unrestricted experimental test bed such that results generalize to human-human interaction, our results, among others (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005), suggest that people do establish basic joint attention also with robots (or artificial agents in general). However, this phenomenon is likely to depend on people's beliefs in the agent's competence or appearance, in particular, when signaling information processes and functionalities different from those of a human.

Acknowledgments

The research reported of in this paper was supported by the IRTG 715 "Language Technology and Cognitive Systems" funded by the German Research Foundation (DFG).

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MA: MIT Press/Bradford Books.

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*, 27-30.

Breazeal, C., Kidd, C., Thomaz, A., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)* (p. 708-713).

Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*, 581-604.

Flom, R., Lee, K., & Muir, D. (Eds.). (2007). *Gaze-Following: Its Development and Significance*. Mahwah, NJ, US: Lawrence Erlbaum Associates.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274-279.

Hanna, J., & Brennan, S. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596-615.

Kreysa, H. (2009). *Coordinating speech-related eye movements between comprehension and production*. Unpublished doctoral dissertation, University of Edinburgh.

Langton, S. R., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Science*, *4*, 50-59.

Meyer, A., Sleiderink, A., & Levelt, W. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25-B33.

Schroeder, M., & Trouvain, J. (2001). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. In *4th isca workshop on speech synthesis.* Blair Atholl, Scotland.

Staudte, M., & Crocker, M. W. (2009a). The effect of robot gaze on processing robot utterances. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society*. Amsterdam, Netherlands.

Staudte, M., & Crocker, M. W. (2009b). Visual Attention in Spoken Human-Robot Interaction. In *Proceedings of the 4th ACM/IEEE Conference on Human-Robot Interaction (HRI'09).* San Diego, USA.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.