In: Kokinov, B., Karmiloff-Smith, A., Nersessian, N. J. (eds.) ***European Perspectives on Cognitive Science***.

ISBN 978-954-535-660-5

Peripheral speaker gaze facilitates spoken language comprehension: Syntactic structuring and thematic role assignment in German

Helene Kreysa (hkreysa@cit-ec.uni-bielefeld.de)

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)
Cognitive Interaction Technology Excellence Cluster
Bielefeld University, 33615 Bielefeld, Germany

Abstract

A speaker's gaze reliably precedes her reference to an object. This consistent link can permit a listener who frontally faces the speaker to inspect the upcoming referent before its mention. It is unclear whether this listener benefit is pervasive even across different sentence structures and when speaker and listener don't face each other directly. A video-taped speaker seated at an angle relative to the camera referred to two out of three characters on a computer screen, using either a subject-verb-object or a non-canonical object-verb-subject sentence. She shifted gaze once from the pre-verbal to the post-verbal referent, a behavior which could, in principle, allow the listener to anticipate which character would be mentioned post-verbally. We recorded eye movements to the characters during spoken comprehension, as well as post-sentence verification times on whether a subsequent schematic depiction of the sentence content correctly depicted the referents (Experiment 1) or the thematic role relations (Experiment 2). The availability of speaker gaze (vs. when no speaker was present) did not affect verification times in either experiment. In the online eye movement measures, however, it led to earlier anticipation of the post-verbal referent, even though the speaker was hardly ever inspected during comprehension. The effect of gaze on fixation patterns interacted with sentence structure, suggesting differential effects of the gaze cue depending on the syntax of the corresponding sentence. Our findings show that (a) speaker gaze effects on listener comprehension (as reflected in eye movements) are robust even when the speaker doesn't frontally face the listener; (b) gaze doesn't just modulate referential processing, but also interacts rapidly with ongoing syntactic structuring and thematic role assignment.

Keywords: spoken sentence comprehension; peripheral speaker gaze; syntactic structuring

Introduction

During spoken language comprehension, a listener's visual attention to objects can be rapidly influenced by information in visual context. Much existing research has focused on how visible objects and depicted events affect visual attention and spoken language comprehension (Altmann, 2004; Chambers, Tanenhaus, Filip, & Carlson, 2002; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Knoeferle, Habets, Crocker, & Münte, 2008; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In these 'visual world' studies, participants hear a sentence while either inspecting an arrangement of real-world objects in front of them, or while viewing a computer display of semi-realistic clipart objects and depicted actions. The speaker of the sentence is usually not visible. In contrast, the visual context of many everyday utterances will often contain the speaker as an integral part. In addition, seeing the speaker may be informative for the listener: When a speaker describes entities in the visual world, s/he robustly gazes at an object before mentioning it (Griffin & Bock, 2000). These pre-mention referent-directed looks of a speaker could, in principle, be exploited by the listener.

Some studies have examined effects of speaker gaze by overlaying a moving cursor on a display, thus representing the speaker's gaze to objects without actually showing the speaker (e.g., Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; Kreysa, 2009). Other experiments have included a real or video-taped speaker (e.g., Hanna & Brennan, 2008; Nappa & Arnold, 2009; Nappa, Wessel, McEldoon, Gleitman, & Trueswell, 2009). In a setup similar to Figure 1, using a collaborative task between two real partners, Hanna and Brennan (2008) showed that seeing the speaker attend to the object she was about to mention led listeners to shift attention to the corresponding object in their own workspace, before the speaker mentioned it. In fact, listeners can even use the gaze of a robot speaker to anticipate the mention of a linguistically ambiguous referent (Staudte & Crocker, 2009). Speaker gaze[1] thus allows listeners to anticipate what a speaker will refer to next (see also Nappa & Arnold, 2009; Staudte & Crocker, 2009, 2010).

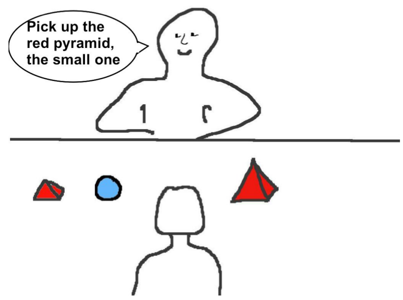


Figure 1: Example for a fully frontal experimental setup.

Gaze effects: Speaker-listener viewing angle and peripheral gaze detection

In all of the studies with a visually present speaker, the speaker faced the listener fully frontally (see Figure 1). Real-life situations in which speaker and listener are fully frontal are at a sales counter, in frontal classroom teaching, or during a lecture. In many other situations, however – e.g., browsing

[1]The term *eye gaze* is used in a wide sense, as a cue to the direction of visual attention. In many cases, this will include head movements.

a shop window with a friend or discussing a shared piece of work with a colleague – we don't have a fully frontal view of each other's eyes. Although gaze direction can be robustly detected when a person faces you, the precision of gaze detection decreases to the extent that the angle (relative to a zero angle for fully frontal view) increases (e.g., Cline, 1967; Gibson & Pick, 1963). A 90° angle between speaker and listener means that only one of the speaker's eyes is visible for the listener, making it more difficult to detect a gaze shift.

It is possible that speaker gaze affects a listener's visual attention only when the listener has a fully frontal view of the speaker (i.e., both eyes are visible and head movements can easily be detected). Alternatively, listeners may be able to detect which objects the speaker shifts her gaze to even when she is positioned at an angle. In this case, a potentially important factor in rapidly detecting (and thus exploiting) such gaze shifts would be the listener's ability to process peripheral information. Overall, it is an open issue whether the findings by Hanna and Brennan (2008) generalize to situations when the speaker doesn't face the listener frontally, but is positioned at an angle (e.g., at 45-60°, see Figure 2a).

Variation of speaker gaze effects by sentence type?

Another open question is whether speaker-gaze-based referent expectation occurs across the board – i.e., independent of the linguistic input – or whether it is sensitive to specific comprehension processes such as syntactic structuring and thematic role assignment. Consider object-subject versus subject-object ordering. In German, both object- and subject-initial main clauses are grammatical, but the latter are canonical while the former are not. Findings from several studies show that unambiguous German object-subject (vs. subject-object) sentences take longer to read (e.g., Hemforth, 1993; Knoeferle & Crocker, 2009).

During spoken comprehension, people can begin to anticipate the post-verbal referent of subject-verb-object (SVO) sentences while hearing the verb. In contrast, when listening to both locally structurally ambiguous (Knoeferle et al., 2005; Weber, Grice, & Crocker, 2006) and unambiguous (Kamide, Scheepers, & Altmann, 2003) object-verb-subject (OVS) sentences, participants initially incorrectly anticipated an object noun referent in post-verbal position, rather than a subject noun referent. Only subsequently, when case marking and world knowledge (Kamide et al., 2003), intonation (Weber et al., 2006), or depicted events (Knoeferle et al., 2005) disambiguated, did people shift attention to the correct post-verbal (subject) referent of the OVS sentences. Processing non-canonical object-initial sentences thus seems to slow comprehension (as reflected by eye movements) relative to subject-initial sentences, suggesting higher overall cognitive load for these sentences. At the same time, case marking, intonation, world knowledge, and visual context factors such as depicted events can modulate this time course. Two experiments investigated whether speaker gaze can function similarly to these other information sources in modulating incremental syntactic structuring and thematic role assignment.

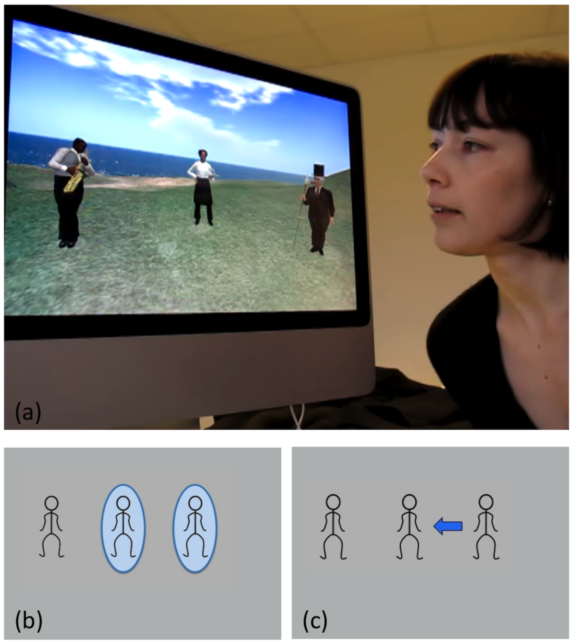


Figure 2: (a) Example of a still from the videos used in Experiments 1 and 2; (b) Experiment 1: Template for reference verification; (c) Experiment 2: Template for verifying role relations.

Motivation and Expectations

The present eyetracking studies examine (a) whether listeners can rapidly use speaker gaze to anticipate a referent before its mention, even from an angle; (b) whether any observed speaker-gaze-based referent anticipation is modulated by syntactic structuring. Knowing more about speaker gaze effects in all sorts of speaker-listener arrangements and across different sentence types and ambiguities will help to extend existing processing accounts of situated comprehension (e.g., Knoeferle & Crocker, 2006) with a speaker model, as well as to inform the construction of human-computer interfaces (e.g., Kopp, Jung, Lessmann, & Wachsmuth, 2003; Poggi, Pelachaud, & Rosis, 2000).

Consider an example: A speaker looks at a computer display that shows a waiter, a millionaire, and a saxophone player (Figure 2a). As soon as the speaker begins her sentence with *Der/ Den Kellner* ("the waiter"), case marking identifies the first noun phrase (NP1) as either the subject (*Der,* Table 1, c & d) or the object (*Den,* Table 1, a & b). Yet when the sentence continues with the verb *beglückwünscht* ("congratulate") and the NP2 determiner *den/ der*, neither linguistic information nor world knowledge can reveal which of the two other depicted characters (the millionaire or the saxophone player) will be referred to post-verbally. Thus, while the sentence is structurally unambiguous, there is a temporary referential ambiguity at the verb.

A speaker's gaze shift to the post-verbal referent could, in principle, identify this "target" referent early, i.e. before hearing the NP2. It is plausible, however, that a listener can't use such shifts of attention immediately, especially in a setup where accurately perceiving gaze direction may be difficult. Also, even if the listener *can* detect the speaker's gaze shift peripherally, it may be that he has to at least briefly inspect the speaker as well, a behavior which – if observed – could lead to temporal delays in any effects of speaker gaze, such that the early referent anticipation observed in prior research is eliminated.

Alternatively, if speaker eye gaze in a setup such as Figure 2a *is* used to anticipate post-verbal referents before their mention, listeners should begin to anticipate this referent shortly after the speaker begins to look at it, and more often than when the display doesn't show the speaker. Such speaker-gaze based anticipation could either be independent of, or modulated by, syntactic structuring and thematic interpretation. If it is independent of sentence structuring, then post-verbal referent anticipation should occur to the same extent and with the same time course for both SVO and OVS sentences. In contrast, if speaker gaze effects on referent anticipation interact with syntactic structuring, then we should see differences in the time course and/ or extent to which a listener inspects the post-verbal referent for OVS relative to SVO sentences. If listeners can exploit speaker gaze better for subject- (vs. object-)initial sentences, we should find an earlier rise of fixations to the target referent when the speaker is present (vs. not) for SVO sentences only. It is possible that with canonical sentences, listeners have ample free processing resources and can thus rapidly make use of gaze shifts, under the assumption that resources are more heavily taxed by the less frequent object-initial sentences. On the other hand, the processing difficulty for OVS sentences could also result in listeners – to the extent that they have free resources – integrating speaker gaze rapidly, precisely because it can help to constrain referential expectations, thus potentially alleviating processing difficulty.

Both experiments used the setup presented in Figure 2. In Experiment 1, participants were asked to verify the sentence referents (see Procedure section). This referential task served to replicate the eye movement findings by Hanna and Brennan (2008) in a situation where the listener is seated at an angle to the speaker. In Experiment 2, participants' task was to verify the thematic relations of the sentence.

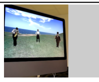
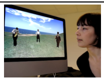
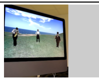
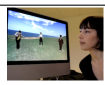
Experiments 1 and 2

Methods

Participants Thirty-two students at Bielefeld University (8 male; 1 replacement) took part in Experiment 1, and a further 32 participated in Experiment 2 (5 male; 3 replacements). All were native German speakers with normal or corrected-to-normal vision, and all gave informed consent.

Materials and Design We created 72 characters using the virtual world Second Life, and 48 critical sentences (NP1-VERB-NP2-PP). We grouped these characters into 24 triplets, and took a snapshot of them. Each of these snapshots was paired with two German sentences (SVO and OVS) to create 24 items. Each sentence described a transitive action

Table 1: Example of the gaze and syntactic structure conditions (congruency is excluded here in the interest of space). The OVS sentence indicates that the waiter is being congratulated by the millionaire, the SVO sentence that he is doing the congratulating.

Condition	**Picture**		**Sentence**
OVS & NoGaze	a		*Den Kellner beglückwünscht der Millionär.*
OVS & Gaze	b		*Den Kellner beglückwünscht der Millionär.*
SVO & NoGaze	c		*Der Kellner beglückwünscht den Millionär.*
SVO & Gaze	d		*Der Kellner beglückwünscht den Millionär.*

taking place between the central character (e.g., the waiter) and one of the two outer characters (e.g., the millionaire; see Table 1). None of the nouns in the sentence were semantically associated, nor was there a semantic connection with the verb. Actions were not depicted. For each item, the two outer characters were referred to in the second noun phrase (NP2) equally often in agent and in patient role across participants. A naming pretest ensured that all characters were recognizable.

We recorded two videos for each item, showing the speaker producing the sentences about the characters. She was seated to the right of a 20" Apple iMac 8.1 screen, which displayed the Second Life triplet. A Canon PowerShot G10 camera was positioned about 1.5 m from the screen and the speaker, in such a way that both were visible in the recording. Videos began with the speaker looking at the camera and smiling briefly to engage the participants' attention. She then looked at all three characters in a fixed inspection order for about six seconds. This inspection period ensured that participants could establish what a gaze to each of the characters looked like. Finally, her gaze returned to the central character, who was always the referent of the NP1. She then began producing the sentence, which had been read out to her before the recording. The speaker always looked at the character she was mentioning. Thus, shortly after uttering the verb, she shifted her gaze from the NP1 referent to the NP2 referent. A second pretest ensured that people could accurately detect this gaze shift (98% correct; detection latency $M = 498$ ms, $SD = 386$).

The design included three within-subject factors (see Table 1): visibility of the speaker ('Gaze' vs. 'NoGaze'), sentence structure (SVO vs. OVS), and congruency between the sentence content and a post-sentence response template (see Procedure). The display versions and sentence manipulations were allocated such that each sentence role (subject or object) was equally distributed across screen positions over the course of the experiment. In addition, the referent of the NP2 was shown on the same side of the screen equally often, so that the speaker shifted her gaze to the right just as frequently as to the left. The 24 experimental items were supplemented by 48 fillers, using different sentence structures and images. Speaker gaze was available on 50% of filler trials.

Procedure We monitored eye movements using an Eyelink 1000 desktop head-stabilized tracker (SR Research). On every trial, participants listened to the speaker's sentence. On Gaze trials, participants saw the speaker talking about the Second Life characters on the screen. On NoGaze trials, the same video was shown, but the speaker was occluded behind a grey bar. Thus, only the static screen with the three characters was visible (see Table 1, a & c). Immediately following the end of each video, participants saw a template like the ones in Figure 2, b & c. Their task was to press a button depending on whether the template accurately depicted ('yes' vs. 'no') the sentence content. For Experiment 1, the template was designed to represent the two characters who had been mentioned in the sentence. For a video such as Figure 2a and the sentence *Den Kellner beglückwünscht der Millionär* ("The waiter is congratulated by the millionaire", OVS), the correct response to the template in Figure 2b would be 'yes', since the positions of the waiter and the millionaire are circled. For Experiment 2, participants verified whether the arrow on the template correctly characterized who-does-what-to-whom in the sentence. Thus, for the same sentence, Figure 2a followed by Figure 2c would also require a 'yes' response, because the arrow points from the millionaire on the right (the agent of the sentence) to the waiter in the middle, who is receiving the congratulations.

Eye movement analysis For the eye movement analyses, we selected two critical time windows during the video. The first window ("SHIFT") comprised all fixations that began between the speaker's gaze shift and the onset of the NP2. The second time window ("NP2") comprised all fixations starting during the second noun phrase. The x-y coordinates of participants' fixations were assigned to areas of interest: NP1 referent, target (= NP2 referent), competitor (= the non-mentioned character), and the area around the speaker. The main dependent variable was the number of fixations to the target, i.e., the referent of the NP2. Log-linear models were used for the inferential analysis, combining characteristics of a standard cross-tabulation chi-square test with those of ANOVA. These models are adequate for count variables because they neither rely upon parametric assumptions concerning the dependent variable, nor do they require linear independence of factor levels (Howell, 2002). Analyses included the factors *Gaze* (gaze vs. no gaze), *Structure* (SVO vs. OVS) and either participants ($N = 32$) or items ($N = 24$).

Results Experiment 1 (Verifying referents)

Response time results Response times were measured from the onset of the template until participants pressed a button on the response pad (97% accuracy). A 2 x 2 x 2 (Structure x Gaze x Congruency) repeated-measures Anova on log-transformed response times revealed significantly faster responses to matching than mismatching templates ($ps < .001$). However, neither Sentence Structure nor Speaker Gaze had any effect on response times ($ps > .5$).

Eye movement results Figure 3 shows proportions of fixations for the Gaze vs. NoGaze conditions in all areas of interest. It illustrates gaze patterns during the SHIFT time window, since this was the earliest point at which gaze information could, in principle, affect a listener's anticipation of the post-verbal referent. Figure 3 shows that participants still focussed to a large extent on the referent of the NP1, who had just been mentioned. However, in the Gaze condition, fixations to the referent of the NP2 increased, despite the fact that it had not yet been mentioned. Note also that fixations to the speaker were rare, even when she was visible.

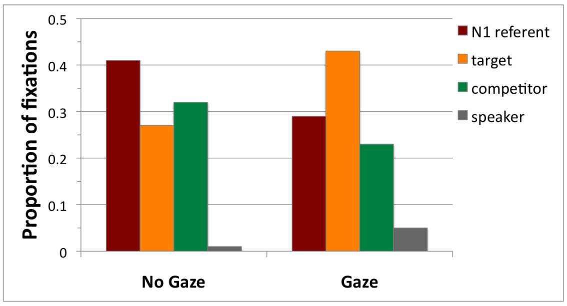


Figure 3: Distribution of fixations beginning in the SHIFT time window across the areas of interest, depending on the visibility of the speaker (Experiment 1).

Figure 4 presents the time course of participants' fixations to the target character only, from the onset of the speaker's gaze shift until three seconds later, as a function of the gaze and sentence structure conditions. Like Figure 3, Figure 4 shows an earlier rise in looks to the target character for both sentence structures in the Gaze (vs. NoGaze) condition. This begins about 500 ms after speaker gaze shift, and well before the onset of the NP2. Only much later, roughly at the offset of the NP2, do participants in the NoGaze conditions fixate the target character to the same extent.

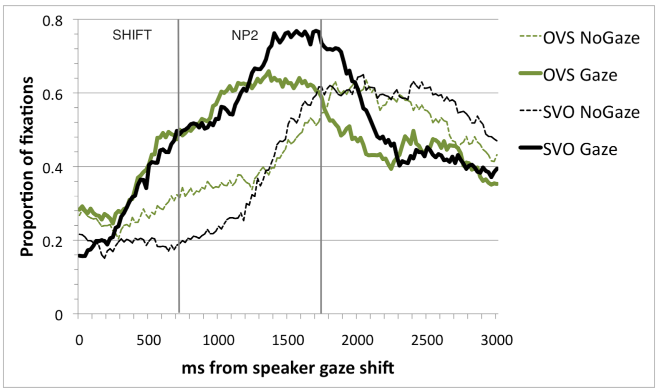


Figure 4: Time course of participants' fixations to the target character (the referent of the NP2) in ms from speaker gaze shift, depending on sentence structure and speaker gaze. The mean on- and offset of the NP2 are marked as vertical lines.

Speaker Gaze affected fixations to the target character in the SHIFT time window: Participants were more likely to fixate the target character when they could see the speaker (43%) than when they could not (27%; $ps < .001$). Gaze and Structure interacted reliably ($ps = .01$): When hearing an OVS sentence, participants fixated the target character 15% more with speaker gaze than without, while this advantage was more than doubled for SVO sentences (38%). In the NP2 time window, the same analysis revealed a main effect of Gaze ($ps < .001$). When the speaker was present, participants fixated the target character more (61%) than in her absence (47%). No other effects were significant in this time window ($ps > .2$).

Results Experiment 2 (Verifying role relations)

Response time results Participants' responses to the who-does-what-to-whom template were 96% accurate. Matching templates again elicited considerably faster responses than mismatches ($ps < .001$). In addition, the role relations verification task (unlike the referential task in Experiment 1) also led to a significant effect of Sentence Structure ($ps < .05$), such that SVO sentences elicited faster responses than OVS (75 ms difference). Again, Speaker Gaze had no reliable effect on response times ($ps > .6$).

Eye movement results Figure 5 shows proportions of fixations in all areas of interest for the Gaze vs. NoGaze conditions during the SHIFT time window. Figure 6 presents the time course of participants' fixations to the target character only, as a function of the gaze and sentence structure conditions. The gaze pattern resembles that for Experiment 1: On trials with a speaker, fixations to the target character began to rise almost as soon as the speaker shifted her gaze, and well before this character was mentioned.

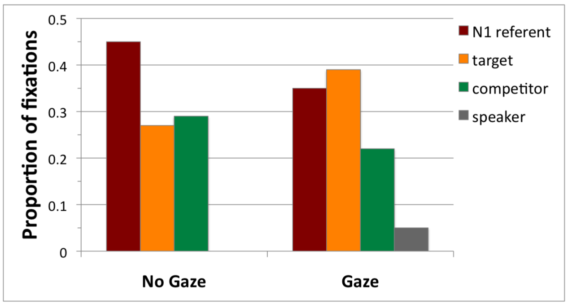


Figure 5: Distribution of fixations in the SHIFT time window across the areas of interest, depending on speaker visibility (Experiment 2).

During the SHIFT time window, log-linear analyses confirmed an effect of Gaze on fixations to the target character: Participants were more likely to fixate the target when they could (vs. couldn't) see the speaker (39% vs. 27%; $ps < .001$). Again, Gaze and Structure interacted ($ps < .05$): When hearing an SVO sentence, participants fixated the target character 30% more with speaker gaze than without, while this advantage was considerably smaller for OVS sentences (10%). Note that the OVS pattern is due in part to a larger proportion of fixations to the NP2 referent in the baseline NoGaze condition, compared to SVO sentences. This result was not predicted, although we believe that it may be caused by a tendency to fixate the agent of the sentence during the verb. We are currently investigating this hypothesis further. As in Experiment 1, in the NP2 time window the only reliable effect was that of Speaker Gaze ($ps < .001$), with participants fixating the target character more often with (63%) than without (47%) gaze.

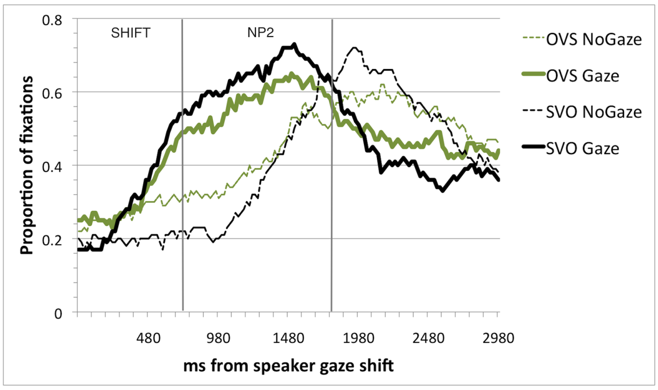


Figure 6: Time course of participants' fixations to the target character, depending on sentence structure and speaker gaze. The mean on- and offset of the NP2 are marked as grey vertical lines.

General Discussion

We examined whether a speaker's gaze shift from a first-mentioned to another referent can rapidly affect listeners' visual attention to that second character before its mention, even when the listener doesn't have the fully frontal view of the speaker that prior studies demonstrating such gaze effects have relied upon. We further assessed whether the listener's visual attention (following the speaker's gaze shift) varies as a function of the syntactic structure of the unfolding utterance. We recorded participants' eye movements as they listened to NP1-VERB-NP2-PP sentences mentioning two out of three virtual characters on a computer screen, as well as verification latencies on whether a post-sentence display indicated the correct NP1 and NP2 characters (Experiment 1), or role relations (Experiment 2).

Response latencies were shorter when the template matched (vs. mismatched) the video in the to-be-verified aspects (Experiments 1 and 2), and shorter for subject- than object-initial sentences (Experiment 2). Whether the speaker was visible or not had no effect on response times. Listeners rarely inspected the speaker during sentence comprehension, but the speaker's gaze shift nonetheless affected the listener's visual attention, as evidenced by an earlier rise of listener fixations to the NP2 referent when the speaker was visible (vs. not). The effect of speaker gaze on listeners' eye movements was modulated by the syntactic structure of the sentence, such that it was more pronounced for subject- than object-initial sentences.

These findings show that (a) speaker gaze can rapidly and peripherally influence visual attention in a listener even when viewed from an angle; (b) speaker gaze effects on visual attention in a listener are not independent of the listener's ongoing syntactic structure building and thematic role assignment; (c) listeners' benefits through speaker gaze were short-lived and not sensitive to modulation through a referential versus thematic role verification task. A processing account of speaker effects on the listener which assumes that speaker gaze interacts purely with referential representations in language comprehension cannot account for the findings of Experiments 1 and 2.

Acknowledgments

This research was funded by the Cognitive Interaction Technology Excellence Center (German research foundation, DFG). We thank Eva Mende, Linda Krull, Anne Kaestner, Lydia Diegmann, and Eva Nunnemann for their assistance with preparing the stimuli and/ or collecting data.

References

Altmann, G. T. M. (2004). Language-mediated eye-movements in the absence of a visual world: the 'blank screen paradigm'. Cognition, 93, B79–B87.

Brennan, S., Chen, X., Dickinson, C., Neider, M., & Zelinsky, G. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. Cognition, 106, 1465–1477.

Chambers, C. G., Tanenhaus, M. K., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real time language comprehension. Journal of Memory and Language, 47, 30–49.

Cline, M. (1967). The perception of where a person is looking. The American Journal of Psychology, 80, 41-50.

Gibson, J., & Pick, A. (1963). Perception of another person's looking behavior. The American Journal of Psychology, 76, 386-394.

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

Hanna, J., & Brennan, S. (2008). Speaker's eye gaze disambiguates referring expressions early during face-to-face conversation. JML, 57, 596–615.

Hemforth, B. (1993). Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens. Sankt Augustin: Infix-Verlag.

Howell, D. C. (2002). Statistical methods for psychology (5th ed.). Pacific Grove: Duxbury.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: cross-linguistic evidence from German and English. Journal of Psycholinguistic Research, 32, 37–55.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. Cognitive Science, 30, 481–529.

Knoeferle, P., & Crocker, M. W. (2009). The time course of constituent order and semantic parallelism: evidence from eye tracking. QJEP, 62, 2338–2371.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. Cognition, 95, 95–127.

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. Cerebral Cortex, 18, 789–795.

Kopp, S., Jung, B., Lessmann, N., & Wachsmuth, I. (2003). Max - a multimodal assistant in virtual reality construction. Künstliche Intelligenz, 4, 11-17.

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

Nappa, R., & Arnold, J. (2009). Paying attention to intention: Effects of intention (but not egocentric attention) on pronoun resolution. In Proceedings of the CUNY Conference (p. 262).

Nappa, R., Wessel, A., McEldoon, K., Gleitman, L., & Trueswell, J. (2009). Use of speaker's gaze and syntax in verb learning. Language Learning and Development, 5, 1–32.

Poggi, C., Pelachaud, F., & Rosis, F. de. (2000). Eye communication in a conversational 3d synthetic agent. The European Journal on Artificial Intelligence (special issue), 13, 169-181.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye-movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. Cognitive Psychology, 45, 447–481.

Staudte, M., & Crocker, M. (2009). The effect of robot gaze on processing robot utterances. In N. Taatgen & H. van Rijn (Eds.), Proceedings of the 31th Annual Meeting of the Cognitive Science Society (pp. 431–436). Cognitive Science Society, Inc.

Staudte, M., & Crocker, M. (2010). When robot gaze helps human listeners: Attentional versus intentional account. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Meeting of the Cognitive Science Society (pp. 1637–1642). Cognitive Science Society, Inc.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268, 632–634.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. Cognition, 99, B63-B72.