Effects of comprehension task and speaker gaze on listeners' allocation of attention in the visual world

Helene Kreysa and Pia Knoeferle
CITEC, Bielefeld University
contact: hkreysa@cit-ec.uni-bielefeld.de

In the past decade, research on spoken language comprehension has been informed by the so-called visual world paradigm, in which comprehenders' eye gaze is recorded while they hear utterances relating to visual context. The assumption is that participants' fixations to objects reflect their incremental utterance comprehension and rapid integration of linguistic and visual information. A further (generally unstated) assumption is that hearing a sentence while viewing a scene generates a strong impulse to map the one onto the other, irrespective of any other goals. The possibility that different tasks (e.g., passive listening versus acting out) might affect this mapping has largely been ignored (Salverda et al., 2011), despite clear evidence that tasks and viewers' moment-to-moment goals can both affect fixation behaviour (Hayhoe & Ballard, 2005). Existing research has further disregarded the possibility that even subtly different comprehension sub-goals (e.g., assigning reference versus thematic roles) might modulate (visual) attention. If this is the case, accounts that aim to model comprehension at the level of these sub-goals need to accommodate their effects (e.g., Knoeferle & Crocker, 2006).

Three eye-tracking experiments presented identical materials but varied a subsequent verification task (each $N = 32$). A videotaped speaker referred to depicted characters, using either a German subject-verb-object (SVO) or a non-canonical object-verb-subject (OVS) sentence. At the verb, the speaker shifted her gaze from the pre-verbal to the post-verbal referent ('gaze' is used in a wide sense here, including head movement). In conditions where the speaker was visible (Fig. 1a, vs. Fig. 1b), this allowed anticipation of the upcoming NP2 referent (cf. Hanna & Brennan, 2007). We compared participants' anticipatory eye movements to this character depending on speaker gaze availability, sentence structure, and verification task: Following the video, participants had to verify whether a schematic depiction correctly highlighted either the two referents (Exp. 1, Fig. 1c), the sentential patient (Exp. 2, Fig. 1d), or the thematic role relations (Exp. 3, Fig. 1e) of the sentence.

Verification response times (RTs) were not affected by speaker gaze but differed between tasks: RTs were faster for reference (Exp. 1) than for patient or role-relations verification (Exps. 2 and 3), and faster to SVO than OVS sentences in role-relations verification only (ps $< .05$).
Eye movements: Speaker gaze rapidly influenced the allocation of attention during comprehension, with earlier fixation of the NP2 referent in all tasks when the speaker was visible than when she was not (ps $< .001$). When verifying the patient (Exp. 2), this anticipation was more pronounced for SVO than OVS sentences (ps $< .01$). This was not the case for referent or role relations verification (Exps. 1 and 3, ps $< .05$), which actually showed somewhat more anticipatory fixations of the referent in OVS than SVO sentences. However, the overall facilitatory effect of speaker gaze was greater in these tasks for subject- than for object-initial sentences (ps $< .05$).

Thus, different post-sentence verification tasks affected not only post-sentence RTs, but also listeners' progressive allocation of visual attention during sentence comprehension. Task affected which character participants inspected, and modulated the effects of both sentence structure and speaker gaze on NP2 anticipation. In sum, visual attention during

comprehension is exquisitely sensitive to response preparation processes, even for subtly differing comprehension subtasks.

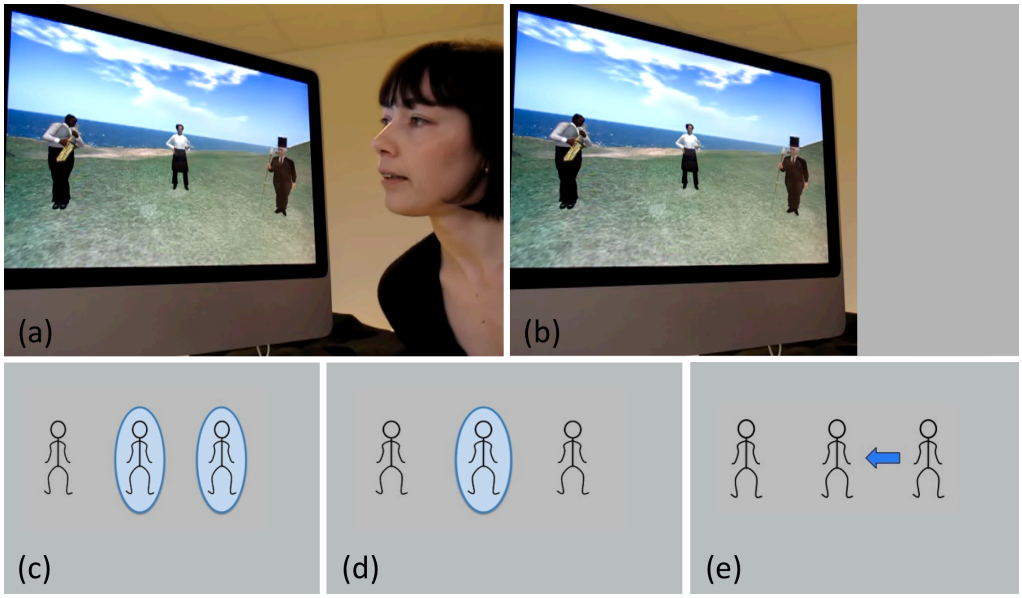

Figure 1: Stills from a comprehension video with (a) speaker visible vs. (b) speaker obscured, and verification templates for the three tasks of verifying (c) the referents (Exp.1), (d) the patient of the action (Exp. 2), and (e) the thematic roles of the sentence (Exp. 3).

References:

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

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*, pp. 188-194.

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

Salverda, A., Brown, M., & Tanenhaus, M. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica, 137*, pp. 172-180.