Congruency effects of speaker's gaze on listeners' sentence comprehension

Torsten Kai Jachmann, Heiner Drenhaus, Maria Staudte, Matthew W. Crocker (Saarland University)
jachmann@coli.uni-saarland.de

We investigate the hypothesis that listeners utilize this external cue as soon as it is available to make predictions about the unfolding sentence, as reflected by neurophysiological indices such as the N400 [1]. Previous eye-tracking studies provided evidence that speaker gaze cues are interpreted by listeners to contain referential intentions [2, 3]. It was shown that participants immediately used the provided gaze cue to disambiguate referring expression, as manifest by a higher inspection rate of the gazed at object compared to a competitor. Additionally, an incongruent gaze cue led to an elevated reaction time when judging sentences for their truth value. However, the underlying nature of processing difficulties of contradictory visual and auditory information remains unclear. We present findings from an ERP study (30 German righthanded participants (age: 18–32)) investigating the influence of speaker's gaze on listeners' understanding of referential expressions in a shared visual scene. In the presented experiment, we utilized a stylized face performing gaze cues time-aligned to an auditory sentence. We manipulated the gaze cue preceding the second noun in the sentence (by 800ms) [4, 5, 6] to investigate the neurophysiological responses to such contradictory information.

Methods: Each experimental item consisted of a visual scene containing three objects that either differed in size (small, medium, large) or brightness (bright, medium, dark) (fully counterbalanced). After three seconds, a stylized face was displayed in the middle of these objects, so that the objects were situated around the face, leaving the position under the face empty to allow for a neutral gaze cue (see Fig 1). Gaze cues were aligned to a spoken comparison of two of the objects of the form "Verglichen mit dem Auto, ist das Haus verhältnismäßig klein, denke ich" ("Compared to the car, the house is proportionally small, I think"). The gaze cue preceding the mentioning of the second noun was manipulated (fully counterbalanced) so as to be: a. congruent (toward the named object); b. incongruent (toward the object that remained unnamed in the sentence) c. neutral (toward the bottom of the screen).

Results: ERPs were time-locked to the start of the second noun following the manipulated gaze cue for the three experimental conditions (Congruent, Incongruent and Neutral). Our analysis revealed a significant difference between the Congruent condition (a) compared to the other two conditions with both conditions (b&c) being globally more negative (at 150ms to 450ms after stimulus onset). Visual inspection further revealed that the Neutral condition (c) contains two, especially frontally distributed, distinct peaks within this time window. A moving time window analysis covering 100ms per time window with a difference of 50ms between each window in this region revealed two globally distributed significant time windows in the Neutral condition compared to the Congruent condition: The early negativity between 150 and 300ms resembles findings from Hagoort and Brown [7], where this early effect is explained as a mismatch between the expected word form given a context and the actual activated word candidates given the speech signal listeners perceive. In our study, the context was introduced visually by the gaze toward an object present in the visual scene (see Connolly and Phillips [8]). Following the early effect, we found a negativity between 300 and 450ms. We interpret this effect as a predictability-driven N400 [1]. In the Neutral condition, two possible referents (the up to this point unnamed objects in the scene) are equally likely to be referred to and hence do not allow for clear predictions. In the Incongruent condition, a clear prediction is made based on gaze that is subsequently violated by the auditory input (noun). In summary, our study highlights that gaze cues can influence the expectancy and predictability of upcoming words in a sentence context and thereby facilitate or impair sentence processing shown in the N2 and the N400 time window. Both effects are modulated by the predictability of the upcoming words on the level of auditory mismatch (N2) and context-driven predictability (N400).

Figures:

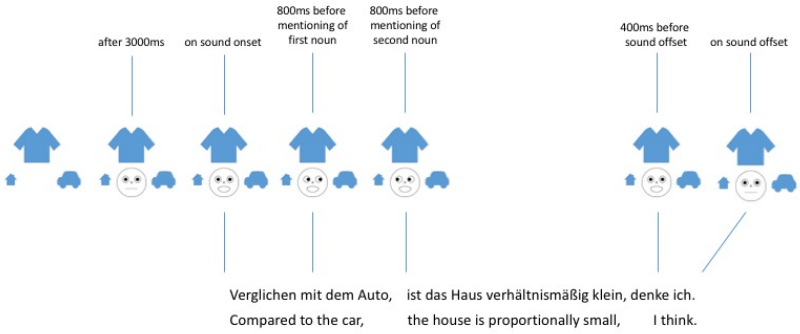

Figure 1: Timeline of a example trial with congruent gaze cues

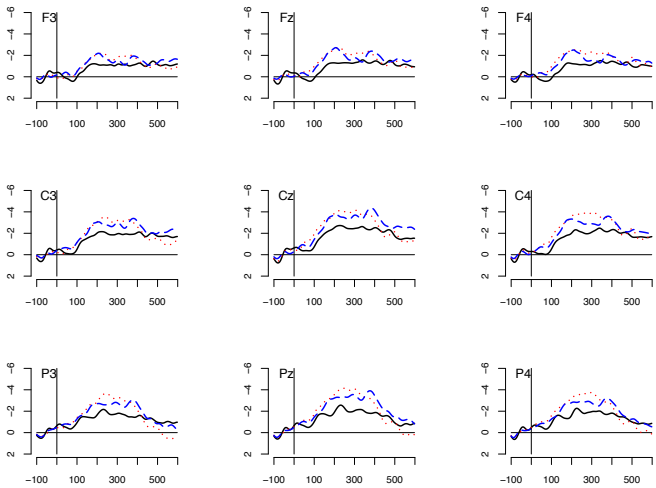

Figure 2: ERP time-locked to the Second Noun Onset separated by the Experimental Conditions (Congruent (solid-black), Incongruent (dotted-red) and Neutral (dashed-blue)). The data presented shows the electrode subset F3, Fz, F4, C3, Cz, C4, P3, Pz and P4.

References:

[1] Dambacher, M., Kliegl, R., Hoffman, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. *Brain Research* 1084, 89–103.

[2] Staudte, M., & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, 120(2), 268–291.

[3] Staudte, M., Crocker, M. W., Heloir, A., & Kipp, M. (2014). The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition*, 133(1), 317–328.

[4] Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition*, 66, 25–33.

[5] Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 4(11), 274–279.

[6] Kreysa, H. (2009). Coordinating speech-related eye movements between comprehension and production. *The University of Edinburgh*.

[7] Hagoort, P., & Brown, C. M. (2000). Erp effects of listening to speech: Semantic erp effects. *Neuropsychologia*, 38(11), 1518–1530.

[8] Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256–266.