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 speech external cue as soon as they are available to make predictions about the unfolding sentence, as reflected by neurophysiological indices such as the N400. 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 right-handed 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) to investigate the neurophysiological responses to such contradictory information.

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 diagonally around the face, leaving one of four position empty. 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 (straight toward the listener).

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 globally distributed significantly larger negativity of both conditions b&c compared to the Congruent condition, resembling findings from Hagoort and Brown (2000), 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. Additionally, an analysis of the time-window from 300-450ms revealed a central-pariatally distributed significantly larger negativity of only the incongruent condition (b) compared to the other two conditions (a&c). We interpret this effect as a predictability-driven N400. In both the Congruent and Neutral condition, the named object is in the set of possible predictions (one for congruent, two for neutral). Only in the incongruent condition, the prediction is violated. Lastly, an analysis of the time-window from 500-1000ms revealed a significantly larger positivity of only the incongruent condition compared to the other two conditions (a&c). We interpret this positivity as evidence of the necessity to update the mental model build based on the previous visual information.