Eyetracking study on the usability of (speaker) gaze when it is associated with the speaker vs when it is not

Description

Here we investigated whether listeners follow and interpret a person's object-directed gaze to convey speaker intentions when those gaze-cues are time aligned like typical speech-aligned speaker gaze cues (SAME condition) EVEN when that person is in fact also listening (and *not* speaking, DIFFERENT condition). For that, we recorded a female's gaze actions to two objects (apple & glass) in front of her but covered her mouth in the video with a grey bar. These videos where coupled with (her) female or another male's utterance of the type "The apple is taller than the glass." Another male person was partially visible in the recording so that the utterances could have plausibly been produced by either person. Lastly, we aligned the gaze actions and utterances in such a way that they were either congruent (gaze to apple, then to glass to match the order of mentioning), incongruent (gaze to glass first, then to apple) or neutral (no object-directed gaze at all) wrt the utterance. Participants were eyetracked as they watched the videos. They further had to evaluate whether the utterances were correct descriptions of the scene or not by button press. The videos were presented in blocks such that listener first saw and heard neutral videos, then videos with congruent gaze, and lastly videos with incongruent gaze. Voice (female/male = same/different) was varied between participants.

Hypothesis

We predicted that it would be more difficult to ignore gaze when plausibly being interpreted as speaker gaze (SAME; female voice). That is, the benefit of such speaker gaze on response times was hypothesized to be equal or higher to co-incidental gaze from another listener. Crucially, we predicted that incongruent speaker gaze (female voice) would be more detrimental and lead to more confusion than co-incidental gaze by a listener (DIFFERENT; male voice). With respect for the blocks, we expected the neutral block to elicit equal response times for both voices. In this block, listener obtained no additional cues from the scene or person and had to rely on language only. In the second block, the congruent gaze and speech was then predicted to elicit faster responses for both voices, over the neutral block, though possibly with different magnitudes.

Result

We found very mixed effects of voice (SAME vs DIFFERENT; i.e. speaker gaze vs coincidental voice) on response times, see graph. The results are difficult to interpret. Especially surprising is the result in the first, neutral block. Here, the male voice elicits slower RTs than the female voice. So there seems to be a bias against the male voice which could have masked other results. In the congruent block, there is a speed up for the male (DIFFERENT) voice, but surprisingly not for the female (SAME) voice. And in the incongruent block, there is a detrimental effect, but contrary to our predictions in the DIFFERENT voice. In fact, it looks like the female voice does not elicit diifferent behavior across conditions, all trials are as fast as the neutral condition trials.

Reasons for this problematic result

We did not see obvious errors in the design. What has been problematic though, is that there seems to be a voice bias, as is obvious from the first, neutral block. We could have tried to counterbalance voice gender to create SAME and DIFFERENT conditions for both voices. However, it would probably not explain why the female voice just did not lead to either speed up nor slow down in combination with gaze.

Data

- Gaze_SameDiff_Stimulus (Gaze_SameDiff.png)
- EyemovementPlots (EyemovementPlots.pdf)
- dissGazeSpeech_newRTs (dissGazeSpeech_newRTs.pdf)