Visual signals as response mobilization cues in face-to-face conversation

Alexandra Emmendorfer[1,2], *Lara Banovac*[3], *Anna Gorter*[4], *Judith Holler*[1,2]

[1]Donders Institute for Brain, Cognition & Behaviour, Radboud University, Nijmegen, The Netherlands

[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[3]Tilburg University, Tilburg, The Netherlands

[4]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

alexandra.emmendorfer@donders.ru.nl; judith.holler@donders.ru.nl

Speakers are remarkably efficient at coordinating turns in face-to-face conversation, with median turn-gaps for question- response sequences between speakers typically falling in the range of 0 – 200 ms relative to question end (Stivers et al., 2009). This poses an intriguing problem for classical psycholinguistic models of language, which propose that at least 600 ms are needed to plan a single word (Indefrey & Levelt, 2004). One characteristic of language use that has long been neglected in psycholinguistics is its multimodality: in face-to-face conversation, the spoken utterance is embedded in a rich multimodal scene including a multitude of communicative visual signals (e.g., Holler & Levinson, 2019). A growing body of evidence suggests that these visual signals may contribute to efficient turn-taking, either by facilitating the identification of an utterance's social action (Nota et al., 2021; Tomasello et al., 2019), allowing the prediction of upcoming words (Ter Bekke et al., 2020), or increasing inter-brain synchrony between interlocuters (Drijvers & Holler, 2022). Speaker gaze is a prominent visual signal in face-to-face communication, however its effect on early response planning is not well understood. We here present data from corpus analyses and an online behavioral study investigating whether speaker gaze may have a response mobilizing effect (Stivers & Rossano, 2010) by increasing the social pressure to respond.

In a corpus of 34 Dutch-speaking dyads, we investigated the prevalence and timing of gaze shifts within question-response sequences, as well as their effect on turn-gap durations. Questions were overall less likely than responses to contain gaze shifts compared to responses, with only 35% of questions (n = 6778) containing gaze shifts, compared to 49% of responses (n = 4553). For both questions and responses containing gaze shifts, the onset of gaze aversions was tightly linked to the onset of the utterance, while the gaze returned to the addressee at variable points throughout the utterance, suggesting that it is an unreliable cue for turn-end estimation. Together, these observations suggest that direct speaker gaze my rather serve a response mobilizing function, signaling to the addressee that a response is expected. If this is the case, we would expect to see shorter turn-gap durations for question-response sequences with static direct gaze, compared to those containing gaze shifts. We restricted our analysis of turn-gap duration to requests for information containing a single gaze shift that did not overlap with unrelated speech. Questions with gaze shifts were classified as *dynamic averted* if the question ended with averted gaze (n = 34), *dynamic direct* if the question ended with direct gaze (n = 170), which were compared to questions with *static direct* gaze (n = 795). Both *dynamic averted* and *dynamic direct* questions led to longer turn-gap durations compared to question with *static direct* gaze (median turn-gaps 695.5 ms, 519 ms and 380 ms, respectively), further supporting the notion of a response-mobilizing role of speaker gaze.

In an online behavioral experiment, we aimed to test these observations experimentally. Participants (data collection ongoing, current n = 32; 29 female, 2 male, 1 non-binary; mean age = 21.5 +/ -3.16) were presented with 240 polar questions (120 general knowledge, 120 personal questions) posed by a virtual avatar, and were instructed to respond as fast and as accurately as possible via button press. The timing of the answer point (i.e., the point at which the listener has enough information to begin planning their response) was manipulated, with half of the questions having an early answer point, the other half a late answer point. We further manipulated the avatar's gaze direction (Figure 1A, 80 trials per condition). The gaze either remained fixed on the participant (*static direct*), or started at an ambiguous point averted by 15deg, before moving toward participant (*dynamic direct*) or further away from participant (30 deg, *dynamic averted*). If speaker gaze acts as a response mobilizer in conversation, we expected questions containing static direct gaze to show faster responses relative to questions containing gaze shifts. We further expected questions ending with averted gaze to show the slowest responses. We further explored whether there would be an interaction between answer point and speaker gaze, where we expected questions with static speaker gaze to show an enhanced effect of answer point. We did not observe any main effect of the presence or direction of gaze shifts on response time (Figure 1B). However, there was an interaction between the presence of gaze shifts and answer point (Figure 1C/D), with questions with gaze shifts showing a larger effect of answer point compared to questions without gaze shifts (β = 137.19 ms, CI = 3.09 – 271.28, p = 0.045). This effect appears to be driven by an increased effect of answer point for *dynamic direct* gaze compared to *static direct* gaze (β = 182.37 ms, CI = 27.13 – 337.61, p = 0.021).

We here report evidence of a role of speaker gaze in response mobilization. While we hypothesized *static direct* speaker gaze would lead to an increased effect of answer point, the current experimental data suggest the strongest effect resulting from dynamic direct gaze. The experimental data is not yet complete so we refrain from interpreting this difference in the results. However, it is important to note that the experiment is void of the wider conversational context present in corpus data, and that the corpus data analyses do not contain answer point as a variable, both of which may affect the results. Future analyses will aim to investigate the influence of these factors. Moreover, one possible interpretation of the current pattern of results may be that listeners are formulating stronger predictions about the unfolding questions with *static direct* gaze, allowing them to already plan their response to questions with late answer points earlier on. In follow up studies, we further investigate the cognitive processes associated with response planning in different gesture and gaze conditions using EEG.

Index Terms: interactive gestures, speaker gaze, response mobilization, turn-taking

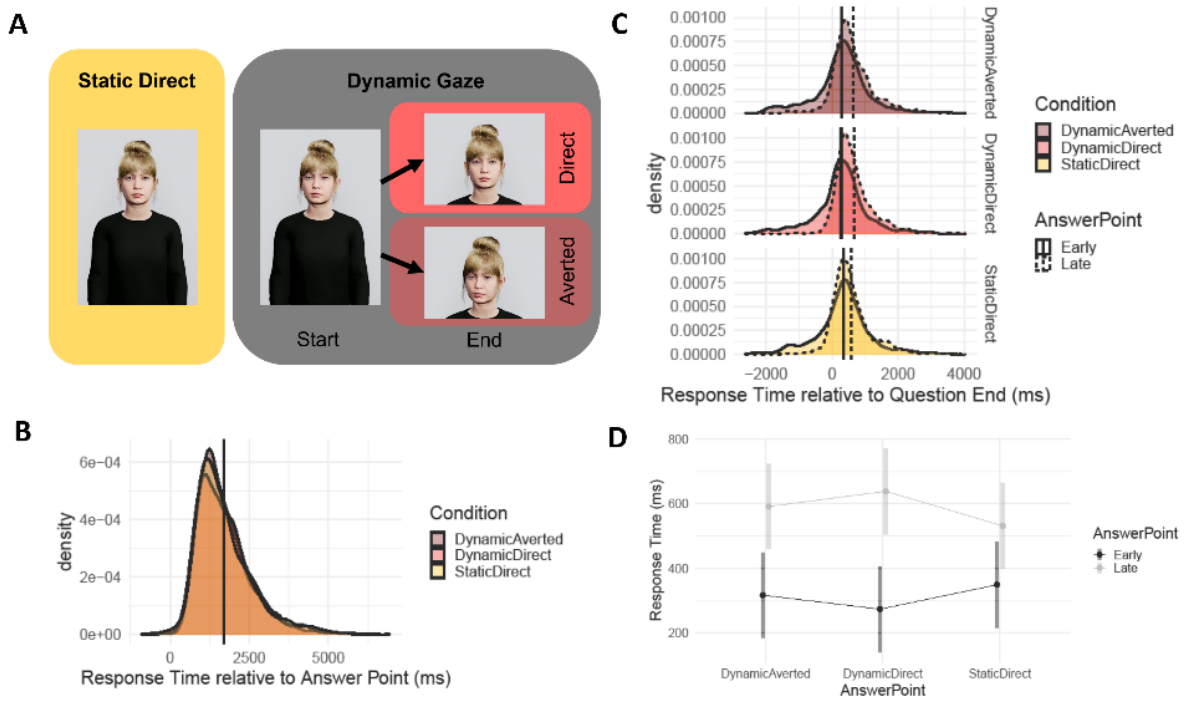

Figure 1: *Experiment 2. (A) Overview of gaze conditions. (B) No effect of gaze condition on response time relative to answer point. (C, D) Larger effect of answer point for questions with gaze shifts, primarily driven by difference between dynamic direct and static direct gaze conditions.*

References

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, *5*(1), Article 1. https://doi.org/10.1038/srep12881

Drijvers, L., & Holler, J. (2022). Face-to-face spatial orientation fine-tunes the brain for neurocognitive processing in conversation. *IScience*, *25*(11), 105413. https://doi.org/10.1016/j.isci.2022.105413

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Nota, N., Trujillo, J. P., & Holler, J. (2021). Facial Signals and Social Actions in Multimodal Face-to-Face Interaction. *Brain Sciences*, *11*(8), Article 8. https://doi.org/10.3390/brainsci11081017

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). *Universals and cultural variation in turn-taking in conversation | PNAS*. https://www.pnas.org/doi/abs/10.1073/pnas.0903616106

Stivers, T., & Rossano, F. (2010). Mobilizing Response. *Research on Language and Social Interaction*, *43*(1), 3–31. https://doi.org/10.1080/08351810903471258

Ter Bekke, M., Drijvers, L., & Holler, J. (2020). *The predictive potential of hand gestures during conversation: An investigation of the timing of gestures in relation to speech*. https://doi.org/10.31234/osf.io/b5zq7

Tomasello, R., Kim, C., Dreyer, F. R., Grisoni, L., & Pulvermüller, F. (2019). Neurophysiological evidence for rapid processing of verbal and gestural information in understanding communicative actions. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-52158-w