Multimodal communication: The impact of an instructor's eye gaze on cognitive performance in a spoken question-answer interaction

Helene Kreysa[1] , Maria Glaser[2], Hanna-Sophia Boltz & Stefan R. Schweinberger[1]

[1]Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University Jena, Germany;
[2]Department of Psychology, Humboldt-Universität zu Berlin, Germany

helene.kreysa@uni-jena.de, maria.glaser@hu-berlin.de, stefan.schweinberger@uni-jena.de

1. Introduction

In multimodal communication, one particular visual cue – **awareness of an interlocutor's gaze behaviour** – uncontroversially affects language comprehension, production, and indeed the entire interaction. However, while several generations of theoretical models exist both for linking face and voice perception [e.g., 1] and for visually situated language comprehension [e.g., 2], their integration remains a challenge. Quite probably, this goal requires further theoretical and methodological advances in so-called "second person neuroscience" [3], but we believe that relevant data can also be contributed through carefully designed experiments in traditional lab settings [e.g., 4] and online.

We present two experiments investigating the impact of a speaker's eye gaze during the questioning stage of a question-answer interaction (Fig. 1) on the cognitive performance of the listener [cf. 6]. Participants were presented with videos of a female speaker either looking at the camera (*direct gaze*) or not (*averted gaze*) while phrasing a problem, which they then had to solve and respond to verbally.

2. Materials and Methods

Experiment 1
- **Participants**: 36 (7 male) (M_{Age} = 21 years)
- **Tasks (blocked presentation)**:

a) forward digit-span [optimal for each individual, range 5-9]
b) verbal problems (synonyms, antonyms, analogies)
c) arithmetic problems

- **Stimuli**: videos of a speaker posing the problem with direct gaze (Fig. 2), averted gaze (Fig. 3), or grey-screen with audio problem only (blocked)
- **Instructions**: speak the answer out loud into a microphone; then press Enter
- **Dependent Variables:** response accuracy and latency (speech onset, speech offset, Enter time)
- **Potential covariates:** Liebowitz Social Anxiety Scale (LSAS [7]), Autism-spectrum Quotient (AQ-k [8])

Experiment 2
- **Participants**: 55 (17 male) (M_{Age} = 22 years)
- **Task**: forward digit-span (DS) : randomized presentation of easy (DS = 5), moderate (DS = 6) and hard (DS = 7) problems
- **Stimuli**: videos from Experiment 1 of the speaker posing the problems with direct gaze (Fig. 1) or averted gaze (Fig. 2)
- **Instructions**: speak the answer out loud into a microphone; then press Enter
- **Dependent Variables:** response accuracy and latency (Enter time)
- **Potential covariates:** LSAS [6], AQ-k [7]).

Fig. 1. Three stages of a question-answer interaction. In the thinking and answering stages, it can be beneficial for the listener to avoid meeting the speaker's gaze [5] to reduce cognitive load. In the questioning stage, however, averted gaze by the speaker may distract attention away from the speaker [6] or signal a lack of interest on their part. This could reduce cognitive performance of the listener.

Fig. 2. Video frame from direct gaze condition.

Fig. 3. Video frame from averted gaze condition.

3. Results

Experiment 1 (Fig. 4)
- no significant differences in performance or response latency depending on presentation mode (direct-gaze, averted-gaze or audio-only)
- no significant interactions with task type (DS, verbal or arithmetic problems)
- faster responses in the audio-only compared to direct gaze conditions of the verbal task for participants reporting higher rates of social anxiety ($p < .05$)
- When rating subjective task difficulty, 34 % of participants found the audio-only condition "most difficult", while 54 % found this condition least difficult.

Experiment 2 (Fig. 5)
- significant main effects for DS difficulty ($F(2, 53) = 209.14$, $p < .001$, $\eta_p^2 = .88$) and Gaze condition ($F(1, 54) = 7.64$, $p = .008$, $\eta_p^2 = .124$): Accuracy in the averted gaze condition ($M = 7.00$, $SD = .37$) was lower than when the speaker appeared to fixate the participant directly ($M = 7.33$, $SD = .35$).
- This pattern was consistent for all three DS difficulties (no significant interaction).
- Participants with higher levels of autistic traits were less accurate for difficult problems presented with direct gaze than participants with lower AQ-k scores.

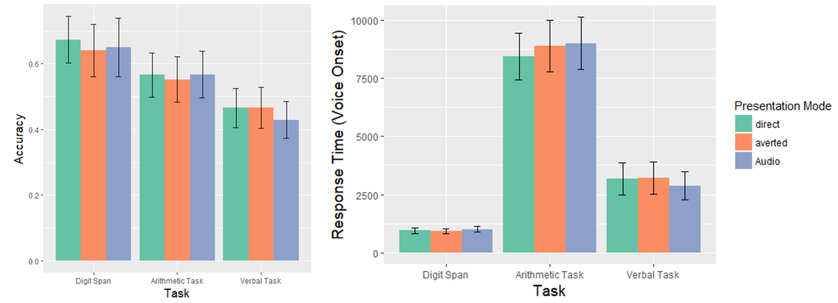

Fig. 4. Experiment 1. Mean accuracies and response latencies across tasks and gaze conditions.

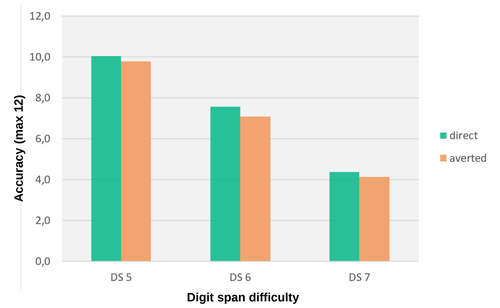

Fig. 5. Experiment 2. Mean accuracies across digit spans and gaze conditions.

4. Conclusions

Follow-up studies are planned using eyetracking to establish whether and at which time-points listeners either seek or avoid meeting the speaker's direct gaze, and how this affects their performance. In addition, Experiments 1 and 2 were conducted before the current pandemic. Since that time most people have had to become much more familiar with video-mediated conversation, including the fact that the necessary technology has rendered direct gaze, which humans are exquisitely tuned to detect in face-to-face contexts [9], much more difficult to identify: In fact, an interlocutor is most likely to appear to be fixating one directly when they are actually looking directly at the web cam rather than into one's eye region [10]. Consequently, we aim to replicate Experiment 2 both in the lab and online in order to address the role of experience for effects of speaker gaze. Finally, we will discuss how these studies can contribute to the larger goal of understanding how visual cues and linguistic content are integrated in multimodal real-life communication.

References

[1] Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and Voice Perception: Understanding Commonalities and Differences. *Trends in Cognitive Sciences, 24*(5), 398-410.
[2] Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language, 112*(3), 189-201.
[3] Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience, 20*(8), 495-505.
[4] Kreysa, H., Nunnemann, E. M., & Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. *Acta Psychologica, 188,* 220-229.
[5] Doherty-Sneddon, G. & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition, 33*(4), 727-733.
[6] Falck-Ytter, T., Carlström, C., & Johansson, M. (2015). Eye contact modulates cognitive processing differently in children with autism. *Child Development, 86*(1), 37-47.
[7] Stangier, U., & Heidenreich, T. (2005). LSAS Liebowitz Social Anxiety Scale. In Collegium Internationale Psychiatriae Scalarum (Ed.), *Internationale Skalen für Psychiatrie* (pp. 299–306). Göttingen: Beltz Test.
[8] Freitag, C. M., et al. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *36*(4), 280–289.
[9] Gibson, J. J., & Pick, A. D. (1963). Perception of Another Person's Looking Behavior. *The American Journal of Psychology, 76*(3), 386-394.
[10] Bailenson, J. N. (2021). Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior, 2*(1).