

Gestures speed up responses to questions

Marlijn ter Bekke, Linda Drijvers & Judith Holler

To cite this article: Marlijn ter Bekke, Linda Drijvers & Judith Holler (17 Feb 2024): Gestures speed up responses to questions, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2024.2314021](https://doi.org/10.1080/23273798.2024.2314021)

To link to this article: <https://doi.org/10.1080/23273798.2024.2314021>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 17 Feb 2024.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

Gestures speed up responses to questions

Marlijn ter Bekke^{a,b}, Linda Drijvers^{a,b} and Judith Holler^{a,b}

^aDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands; ^bMax Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ABSTRACT

Most language use occurs in face-to-face conversation, which involves rapid turn-taking. Seeing communicative bodily signals in addition to hearing speech may facilitate such fast responding. We tested whether this holds for co-speech hand gestures by investigating whether these gestures speed up button press responses to questions. Sixty native speakers of Dutch viewed videos in which an actress asked yes/no-questions, either with or without a corresponding iconic hand gesture. Participants answered the questions as quickly and accurately as possible via button press. Gestures did not impact response accuracy, but crucially, gestures sped up responses, suggesting that response planning may be finished earlier when gestures are seen. How much gestures sped up responses was not related to their timing in the question or their timing with respect to the corresponding information in speech. Overall, these results are in line with the idea that multimodality may facilitate fast responding during face-to-face conversation.

ARTICLE HISTORY

Received 30 March 2023
Accepted 21 January 2024

KEYWORDS

Multimodal language;
gesture; turn-taking;
conversation; facilitation;
response times

Introduction



In face-to-face conversations, people rapidly take turns at talking: responses to questions typically occur within 0–200 ms after question end (Stivers et al., 2009). Slower response times can mark responses as dis-preferred (Kendrick & Torreira, 2015; Schegloff, 2007) or indicate weaker social connection (Templeton et al., 2022). Thus, response times in conversation matter, and therefore fast language processing and response planning are of the essence.


During conversational turn-taking people also use bodily signals to communicate (Bavelas, 2022; Kendon, 2004). These signals may facilitate fast language processing (Drijvers & Holler, 2022; Holler & Levinson, 2019) and thus allow for earlier responses. We tested this hypothesis by investigating whether iconic co-speech hand gestures speed up responses to questions.

Iconic hand gestures depict semantic information about concrete referents like actions or objects (McNeill, 1992), e.g. depicting drinking by bringing a hand to the mouth, as if holding a glass. These gestures could facilitate fast responding in several (not mutually exclusive) ways. Responding in conversation requires comprehension of the incoming utterance and simultaneous response planning, after which the response is

launched when turn-final cues are detected (Bögels et al., 2015; Levinson, 2016; Levinson & Torreira, 2015). Gestures could facilitate comprehension and response planning by, for example, adding semantic information that helps grasp the utterance's meaning (Dargue et al., 2019; Holler et al., 2018), increasing attention to speech (Dargue et al., 2019; Holler et al., 2018), or improving prediction of upcoming words (Holler & Levinson, 2019; ter Bekke et al., 2024; Zhang et al., 2021). Subsequently, gestures could result in responses being *launched* earlier by gesture retractions signalling upcoming turn ends (Duncan, 1972; Holler et al., 2018; Levinson & Torreira, 2015). Using an experimental button press paradigm which encouraged participants to respond as fast as possible (irrespective of whether the speaker had finished speaking), we focused on the first stage – the possibility that iconic gestures influence when participants *can* respond (i.e. gestures facilitating response planning).

Gestures may possibly not speed up or even slow down response planning. Parallel comprehension and response planning is cognitively taxing. If listeners additionally have to process visual signals and integrate them with speech where relevant, this may significantly strain our processing system and slow down responses (Holler & Levinson, 2019).

CONTACT Marlijn ter Bekke  Marlijn.terBekke@donders.ru.nl  Donders Institute for Brain, Cognition and Behaviour, Radboud University, Thomas van Aquinostraat 4, 6525 GD, Nijmegen, the Netherlands

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2024.2314021>.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Correlational corpus data favour accounts postulating a facilitative role of visual signals: questions with hand gestures get faster responses than questions without (Holler et al., 2018; ter Bekke et al., 2024). However, no causal claims can be made based on corpus data, making it unclear if gestures speed up responses, or a confounding variable causes the effect. For example, questions people gesture with might generally be easier to understand or structured differently, enabling earlier responses.

Controlled experimental studies also reported gestural speed up effects. When participants saw videos with speech and iconic gestures (versus speech-only), they were faster to judge whether the videos matched actions (Kelly et al., 2010) or pictures (Krason et al., 2021; Wu & Coulson, 2015), or whether they felt addressed by the speaker (Nagels et al., 2015). These results align with lower-level multimodal facilitation effects where participants are faster to detect audiovisual stimuli versus audio-only (e.g. Miller, 1986).

However, gestural speed up effects were not always found. Participants did not read words aloud faster when preceded by videos with a pantomime of the word (Bernardis et al., 2008). Similarly, participants did not judge faster whether someone described an object- or person-related event when they also produced emblems or pantomimes (He et al., 2015). In three other studies (Drijvers et al., 2018; Drijvers et al., 2019; Drijvers et al., 2019), seeing an actress produce iconic gestures did not speed up word recognition.

Altogether, the experimental evidence that gestures speed up responses is mixed. Although these studies differed in several ways, their tasks involved participants making judgements (e.g. about videos matching pictures), rather than planning responses that are contingent on previous utterances as in conversation. This makes it difficult to directly link these results to the corpus finding that questions with gestures got faster responses.

Our key aim was to test the gestural speed up effect experimentally, to bring the corpus and experimental findings together. Participants viewed videos in which an actress asked them yes/no-questions (e.g. “is being able to type well useful if you’re a secretary?”) and had to answer as fast and accurately as possible, via button press. The exact same questions were presented either with or without a corresponding iconic gesture (e.g. fingers pretending to type). We opted for responding via button press rather than verbally for several reasons. Firstly, this way we retained control over the complexity and duration of the responses, which are known to affect their timing (Roberts et al., 2015). Secondly, buttons can be easily pressed *during* a question without the norms and principles governing overlap management in conversation coming into play (Sacks

et al., 1974). This gives us a more direct measure of when recipients *can* respond (the focus on the present study), rather than when recipients *would* respond in more conversational settings.

If gestures speed up responses, this would support the idea that multimodal language is easier/faster to process than unimodal language (Drijvers & Holler, 2022; Holler & Levinson, 2019). It would advance the literature by showing that iconic co-speech gestures specifically are one of the signals contributing to multimodal facilitation.

Moreover, visual signals produced earlier in the question may exert greater influence on fast responding (for such effects of eyebrow movements, see Nota et al., 2022). As the gesture timing spontaneously varied across questions (i.e. one gesture may occur earlier than another), we explored whether gestures especially sped up responses when they had earlier preparations, strokes or retractions (Kita et al., 1998), which refer to the hand(s) moving from their resting position into gesture space and preparing the gestural handshape (preparation phase), depicting the action or object (stroke phase), and returning to the resting position (retraction phase). Furthermore, if gestures speed up responses by improving predictions about upcoming words, then gestures that precede their corresponding word in speech more (i.e. more predictive potential) may speed up responses more (ter Bekke et al., 2024).

Methods

The pre-registration is available here: <https://aspredicted.org/dp573.pdf>.

Data, scripts and supplementary materials can be found here: <https://osf.io/e6wpt/>.

Participants

Sixty right-handed native speakers of Dutch (51 women, $M_{\text{age}} = 23;2$ [years;months], $SD = 4;6$) participated. All reported normal hearing, (corrected-to-)normal vision, and no neurological/motoric/language-related disorders (same for pre-test participants). Participants gave informed consent before the experiment and received financial compensation/course credits. The experiment was approved by the Social Sciences Faculty Ethics Committee, Radboud University (approval code: ECSW-2018-135).

Stimuli

Participants watched short videos of a native Dutch actress asking easy yes/no-questions. Questions were

factual to ensure 50% yes-responses and 50% no-responses. Personal questions were included as fillers. For a list of all questions, see <https://osf.io/e6wpt/>.

Videos were recorded with a Canon XF205 camera (50fps) and displayed the actress from head to knees (Figure 1). She was instructed to utter each question as she normally would. In the Gesture condition, she was asked to produce a gesture depicting the lexical affiliate (i.e. the word in the question corresponding to the gesture) in a way that felt natural. All gestures were iconic depictions (e.g. pretending to swim for “swimming”), except one which was therefore excluded, reducing the number of items. A pre-test with 20 different participants confirmed the gestures (without audio) in general fit the lexical affiliates (see Appendix A). Twelve gestures (7 fillers) were not well-understood and excluded.

To test how well, in their question context, the gestures fit their lexical affiliates (iconicity), and how natural they looked, a second pre-test showed the gesture clips *with* audio to another 20 different participants (see Appendix B). Gestures generally fit the lexical affiliates (mean iconicity rating 5.0 on 1–7 scale, range: 3.1–6.8) and were rated as rather natural (mean naturalness rating 4.8 on 1–7 scale, range: 3.2–6.2). As these ratings varied considerably across items, we explored whether they predicted how much each gesture sped up responses.

A skin-coloured square was added in Adobe Premiere Pro to cover visible speech movements from the face/throat (Figure 1). This was done for both conditions, allowing us to use the Gesture condition audio also in the No-gesture condition. Audio was intensity-scaled (70 dB) and denoised in Praat (Boersma, 2001). Because gestures could start before or end after the question, videos were cut from 1000 ms before to 1000 ms after the question. On average, gestures started 188 ms before question onset ($SD = 311$ ms) and ended 999 ms before question offset ($SD = 616$ ms).

Coding

We annotated gesture phases frame-by-frame in ELAN (version 5.5; Wittenburg et al., 2006). Based on the hands being blurry or clear, gestures were segmented into static or dynamic movement phases (Seyfeddinipur, 2006) and identified as preparation, stroke, retraction, or pre- or post-stroke hold (i.e. hands are still before or after the stroke; Kita et al., 1998). Gesture phase coding was reliable (Appendix C).

Procedure

In this online study (Gorilla; Anwyl-Irvine et al., 2020), participants were instructed to participate in a quiet

location, in one go, with wired ear-/headphones and other computer programs closed. A questionnaire ensured eligibility. After a sound check, the main experiment started. Participants were instructed to watch each video and answer the question as quickly and accurately as possible via button press, without waiting until question end. They were not informed about the gesture manipulation.

Each trial started with a 1000 ms fixation cross, followed by the video. Participants responded with a *yes* or *no* button press (button assignment was counterbalanced), after which a new trial started immediately.

After three practice trials per condition, the stimuli were presented in four blocks with self-paced breaks in between. Each participant saw 134 items¹ (32 fillers) in a within-subjects design: 67 in each condition (16 fillers). Within each condition, the amount of *yes*- and *no*-responses was counterbalanced. Across participants each item was seen in both conditions (50% Gesture; 50% No-gesture). Stimulus order was pseudo-randomized per participant, with no more than two repetitions of one condition in a row. Finally, participants were debriefed.²

Analysis

We excluded five responses slower than 5000 ms (0.1% of data; three in the Gesture condition), eight responses before question onset (0.1%; four Gesture), and three cases where the video did not play (one Gesture). For response time analyses, we excluded 276 responses with incorrect answers (4.5%; 129 Gesture).

We fitted (generalised) linear-mixed effects models using lme4 (1.1-26; Bates et al., 2015) in R (4.2.1; R Core Team, 2019) and lmerTest (3.1-1; Kuznetsova et al., 2017). Factors were sum-to-zero contrast coded (–0.5, 0.5). Continuous predictors causing scaling errors were z-standardized. We used maximal random effects structures. In case of convergence issues, we used estimates from the non-converged fit as starting values, and then compared the estimates from different optimisers using allFit(). If these converged to highly similar values, the warnings were considered false positives (RDocumentation, 2021). This occurred for all analyses below.

To test whether Condition (Gesture, No-gesture) affected accuracy (0, 1) and response times (in ms), we used models with Condition as predictor. To test whether gestures with earlier gesture phases (preparation, stroke, retraction) especially sped up responses, we calculated the difference between each gesture phase onset and question offset (in ms), and used this as predictor. This slightly deviates from our pre-registration. First, we tested the effect of Gesture phase

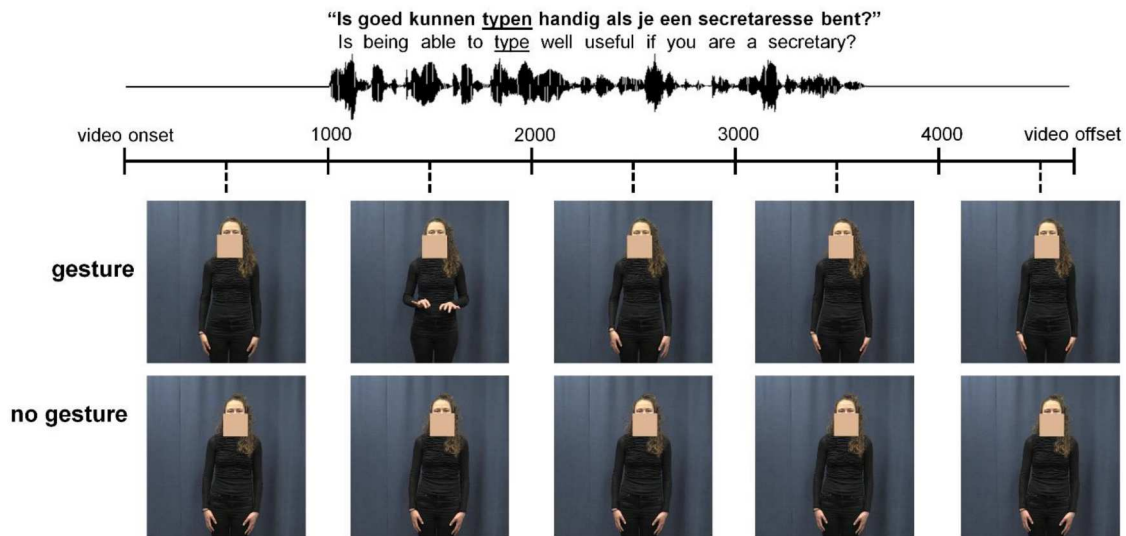


Figure 1. Schematic stimulus overview.

onset in interaction with Condition, not as main effect, because gesture phase onset timing should only matter in the Gesture condition (where participants actually saw this gesture phase). Second, the preparation and stroke onset analyses were added. To explore whether gestures that preceded their lexical affiliate more sped up responses more, we calculated the difference between the preparation/stroke onset and the lexical affiliate onset and used this difference as a predictor of response times.

To explore whether iconicity and naturalness scores predicted how much gestures sped up responses, we used a data-driven model-building procedure, starting with only Condition as predictor and adding interactions stepwise (Iconicity*Condition first). We only kept factors/interactions if they significantly improved model fit.

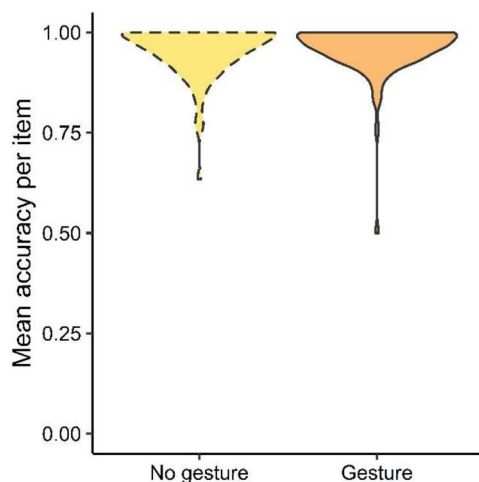


Figure 2. Accuracy did not differ by condition. Distributions of mean accuracy for each item are shown per condition.

Results

Accuracy

The questions were easy to answer, with 95.5% ($SD = 20.8\%$) correct responses (Gesture: 95.8%, $SD = 20.1\%$, No-gesture: 95.2%, $SD = 21.4\%$) (Figure 2). Accuracy did not differ between questions with and without gestures ($\beta = 0.25$, $SE = 0.19$, $z = 1.29$, $p = 0.20$).

Response times

Overall, the mean response time was 160.6 ms ($SD = 529.94$ ms). Without gestures, it was 186.9 ms ($SD = 524.2$ ms) and with gestures around 50 ms faster (134.5 ms, $SD = 534.4$ ms; Figure 3). Indeed, gestures sped up responses to questions³ ($\beta = -52.65$, $SE = 10.81$, $t = -4.87$, $p < 0.01$).

Gesture phase timing

How early gesture phases started in the question did not predict how much gestures sped up responses, for preparations ($\beta = 12.29$, $SE = 10.92$, $t = 1.13$, $p = 0.26$), strokes ($\beta = 11.96$, $SE = 10.87$, $t = 1.10$, $p = 0.27$), nor retractions ($\beta = 14.30$, $SE = 10.96$, $t = 1.31$, $p = 0.20$).

Predictive potential

All preparations started before their lexical affiliate, on average 753 ms before ($SD = 268$ ms). Moreover, most strokes (81%) started before their lexical affiliate, on average 227 ms before ($SD = 225$ ms). How much preparations ($\beta = 0.05$, $SE = 0.04$, $t = 1.35$, $p = 0.18$) or strokes ($\beta = 0.06$, $SE = 0.05$, $t = 1.43$, $p = 0.15$) preceded lexical

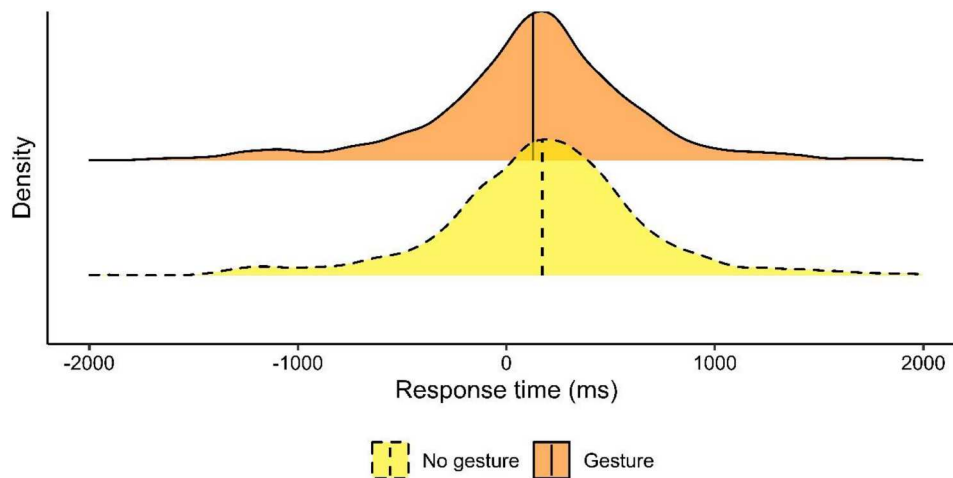


Figure 3. Gestures sped up responses to questions. The response time distribution per condition is shown. Vertical lines display mean response times.

affiliates did not predict how much gestures sped up responses.

Gesture iconicity and naturalness

Iconicity ($\beta = 2.87$, $SE = 11.98$, $t = 0.24$, $p = 0.81$) and naturalness ratings ($\beta = 0.19$, $SE = 16.24$, $t = 0.01$, $p = 0.99$) did not predict how much gestures sped up responses.⁴

Discussion

In face-to-face conversation, people rapidly take turns and use hand gestures to communicate. This study provides the first experimental evidence that these gestures speed up responses to questions, in line with corpus work showing that questions with gestures get faster responses (Holler et al., 2018; ter Bekke et al., 2024) and the *multimodal facilitation hypothesis* stating that communicative bodily signals facilitate fast language processing (Drijvers & Holler, 2022; Holler & Levinson, 2019). Despite limited cognitive capacities and strong time pressure (in conversation, and likewise in our experiment), seeing iconic hand gestures facilitates fast responding.

Using an experimental button press paradigm which encouraged participants to respond as fast as possible, we focused on whether iconic gestures influence when participants *can* respond (i.e. gestures facilitating response planning), rather than when participants would respond in actual conversation, where turn-taking principles and conversational pragmatics influence response timing (Kendrick & Torreira, 2015; Sacks et al., 1974). The present study showed that the perception of gestures led to faster responses. Critically, the perception of gestures did not lead to responses being more closely timed to the question end (see

Supplementary Materials) and gestures did not speed up responses more if retractions (a potential turn-final cue; Duncan, 1972). We thus interpret the results as suggesting that gestures allow for faster response planning, possibly due to faster comprehension. To address the different question of whether gestures may influence when participants *launch* their responses in actual conversation, as gestures can also function as turn-final cues and manage turn transitions in conversation (Duncan, 1972; Holler et al., 2018; Kendrick et al., 2023; Mondada, 2006; Zellers et al., 2019), a paradigm able to simulate the sensitivities surrounding gaps and overlaps in turn-taking would be needed.

The mechanism underlying this gestural facilitation may be gestures providing semantic information (Dargue et al., 2019; Dargue & Sweller, 2018; Holler et al., 2018) or increasing attention to speech (Dargue et al., 2019; Holler et al., 2018). In this study, it appears unlikely that gestures sped up responses by facilitating prediction of upcoming words (Holler & Levinson, 2019; ter Bekke et al., 2024; Zhang et al., 2021). Although most gestures started before their lexical affiliates, similar to natural conversation (Donnellan et al., 2022; ter Bekke et al., 2020; ter Bekke et al., 2024), gestures did not speed up responses more when gestures occurred earlier. However, as gesture timing was not experimentally manipulated, we cannot draw strong conclusions here. Future studies may delve deeper into the underlying mechanisms of gestural facilitation, for example by varying the task or stimuli (e.g. additional control conditions to contrast possible mechanisms), or by looking at individual differences and how these relate to the underlying mechanisms (Özer & Göksun, 2020).

Interestingly, our gestural speed up effect was smaller (~50 ms faster) than in past corpus studies (~300 ms

faster). Perhaps gestures are more effective in an interactive, social context. Thus, our experimental task, with video stimuli rather than a live interlocutor, and button press responses rather than verbalised responses, may have reduced gestures' effect. Alternatively, it is possible that the effect size attributed to gestures in the corpus studies was partly caused by confounding variables. Many factors influence response times (e.g. question duration; Roberts et al., 2015), and one of these may have correlated with gesture presence. This highlights the importance of complementing corpus studies with controlled experiments that use identical questions with and without gesture, but also the need for building on the present findings with more ecologically valid, interactionally-embedded experimental paradigms. Such studies could, for example, manipulate gesture presence in the context of other cues (e.g. visible speech, gaze), let participants respond verbally, or add interactivity or co-presence. Overall, converging evidence from approaches varying from more ecologically valid to more experimentally controlled seems best suited to uncover robust patterns in multimodal language processing.

In conclusion, we demonstrated that iconic gestures speed up responses to questions. While previously found in corpora, our experimental evidence is the first to showing iconic gestures' causal role in fast responding. Despite the challenge of processing and integrating different signals under significant time constraints, multimodality may actually facilitate fast responding.

Notes

1. Two items were excluded to ensure proper counterbalancing.
2. Participants' Empathy Quotients were also collected but did not associate with how much participants benefited from seeing the gesture. For these supplementary analyses, see <https://osf.io/e6wpt/>.
3. Gestures did not make response times more precise (i.e. closer to question end), but rather sped up responses in general, making overlap responses less precise and gap responses more precise on average. For these supplementary analyses, see <https://osf.io/e6wpt/>.
4. Iconicity also did not modulate the relation between gesture phase timing (with respect to question offset or lexical affiliate onset) and how much gestures sped up responses. For these supplementary analyses, see <https://osf.io/e6wpt/>.

Acknowledgements

We are grateful to Lina van Otterdijk for help with reliability coding, Jeroen Geerts for help with video editing, Maarten van den Heuvel for Presentation help and the actress for her help producing the stimuli.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

J.H. and M.t.B. were funded by the European Research Council (CoG grant #773079, awarded to J.H.). L.D. was funded by the Max-Planck-Gesellschaft (Minerva Fast Track Fellowship)

References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bavelas, J. B. (2022). *Face-to-face dialogue: Theory, research, and applications*. Oxford University Press.
- Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioural and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, 25(7-8), 1114–1128. <https://doi.org/10.1080/02643290801921707>
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9-10), 341–345.
- 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), 12881. <https://doi.org/10.1038/srep12881>
- Dargue, N., & Sweller, N. (2018). Not all gestures are created equal: The effects of typical and atypical iconic gestures on narrative comprehension. *Journal of Nonverbal Behavior*, 42(3), 327–345. <https://doi.org/10.1007/s10919-018-0278-3>
- Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, 145(8), 765–784. <https://doi.org/10.1037/bul0000202>
- Donnellan, E., Özder, L. E., Man, H., Grzyb, B., Gu, Y., & Vigliocco, G. (2022). *Timing relationships between representational gestures and speech: A corpus based investigation*. Proceedings of the Annual Meeting of the Cognitive Science Society, 44. <https://escholarship.org/uc/item/7w349725>.
- Drijvers, L., & Holler, J. (2022). The multimodal facilitation effect in human communication. *Psychonomic Bulletin & Review*, 30(2), 792–801. <https://doi.org/10.3758/s13423-022-02178-x>
- Drijvers, L., Özyürek, A., & Jensen, O. (2018). Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Human Brain Mapping*, 39(5), 2075–2087. <https://doi.org/10.1002/hbm.23987>
- Drijvers, L., Vaitonytė, J., & Özyürek, A. (2019a). Degree of language experience modulates visual attention to visible speech and iconic gestures during clear and degraded speech comprehension. *Cognitive Science*, 43(10), e12789. <https://doi.org/10.1111/cogs.12789>

- Drijvers, L., van der Plas, M., Özyürek, A., & Jensen, O. (2019b). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *NeuroImage*, *194*, 55–67. <https://doi.org/10.1016/j.neuroimage.2019.03.032>
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292. <https://doi.org/10.1037/h0033031>
- Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology*, *23*(2), 204–214. <https://doi.org/10.1177/0261927X04263825>
- He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., & Straube, B. (2015). The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia*, *72*, 27–42. <https://doi.org/10.1016/j.neuropsychologia.2015.04.018>
- Holle, H., & Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behavior Research Methods*, *47*(3), 837–847. <https://doi.org/10.3758/s13428-014-0506-7>
- Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, *25*(5), 1900–1908. <https://doi.org/10.3758/s13423-017-1363-z>
- 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>
- Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. <https://doi.org/10.1177/0956797609357327>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511807572>
- Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), 20210473. <https://doi.org/10.1098/rstb.2021.0473>
- Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes*, *52*(4), 255–289. <https://doi.org/10.1080/0163853X.2014.955997>
- Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction* (Vol. 1371) (pp. 23–35). Springer. <https://doi.org/10.1007/BFb0052986>
- Krason, A., Fenton, R., Varley, R., & Vigliocco, G. (2021). The role of iconic gestures and mouth movements in face-to-face communication. *Psychonomic Bulletin & Review*, *29*(2), 600–612. <https://doi.org/10.3758/s13423-021-02009-5>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. <https://doi.org/10.2307/2529310>
- Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. <https://doi.org/10.1016/j.tics.2015.10.010>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731. <https://doi.org/10.3389/fpsyg.2015.00731>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343. <https://doi.org/10.3758/BF03203025>
- Mondada, L. (2006). Participants' online analysis and multimodal practices: Projecting the end of the turn and the closing of the sequence. *Discourse Studies*, *8*(1), 117–129. <https://doi.org/10.1177/1461445606059561>
- Nagels, A., Kircher, T., Steines, M., & Straube, B. (2015). Feeling addressed! The role of body orientation and co-speech gesture in social communication. *Human Brain Mapping*, *36*(5), 1925–1936. <https://doi.org/10.1002/hbm.22746>
- Nota, N., Trujillo, J., & Holler, J. (2022). *Conversational eyebrow frowns facilitate question identification: An online VR study*. PsyArXiv. <https://doi.org/10.31234/osf.io/fcj8b>
- Özer, D., & Göksun, T. (2020). Gesture use and processing: A review on individual differences in cognitive resources. *Frontiers in Psychology*, *11*, 573555. <https://doi.org/10.3389/fpsyg.2020.573555>
- Qualtrics. (2021). [Computer software]. <https://qualtrics.com>.
- R Core Team. (2019). R: A Language and Environment for Statistical Computing [Computer software]. <https://www.R-project.org/>.
- RDocumentation. (2021). *Convergence: Assessing convergence for fitted models*. <https://www.rdocumentation.org/packages/lme4/versions/1.1-27.1/topics/convergence>.
- Roberts, S. G., Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: A corpus study. *Frontiers in Psychology*, *6*, 509. <https://doi.org/10.3389/fpsyg.2015.00509>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. <https://doi.org/10.1353/lan.1974.0010>
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge University Press.
- Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture* Unpublished doctoral dissertation, Radboud University. https://pure.mpg.de/rest/items/item_59337/component/file_2291735/content.
- 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. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings*

- of the *National Academy of Sciences*, 119(4), e2116915119. <https://doi.org/10.1073/pnas.2116915119>
- 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*. PsyArXiv. <https://doi.org/10.31234/osf.io/b5zq7>.
- ter Bekke, M., Drijvers, L., & Holler, J. (2024). Hand gestures have predictive potential during conversation: An investigation of the timing of gestures in relation to speech. *Cognitive Science*, 48(1), e13407. <https://doi.org/10.1111/cogs.13407>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *Proceedings of the 5th International Conference on Language Resources and Evaluation*, 1556–1559.
- Wu, Y. C., & Coulson, S. (2015). Iconic gestures facilitate discourse comprehension in individuals with superior immediate memory for body configurations. *Psychological Science*, 26(11), 1717–1727. <https://doi.org/10.1177/0956797615597671>
- Zellers, M., Gorisch, J., House, D., & Peters, B. (2019). Hand gestures and pitch contours and their distribution at possible speaker change locations: A first investigation. *Proceedings of Gesture and Speech in Interaction*, 93–98.
- Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2021). More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proceedings of the Royal Society B: Biological Sciences*, 288(1955), 20210500. <https://doi.org/10.1098/rspb.2021.0500>