Audience effects and production demands on timing relationships between representational gestures and speech

Ed Donnellan[*1], Yumeng Wang*[*1], Levent Emir Özder[2], Hillarie Man[1], Kellie Fraser[3], Amara Jiménez Cañizares[1], Beata Grzyb[1], Yan Gu[4], Gabriella Vigliocco[1]*

[1]University College London, [2]Koç University, [3]University of Aberdeen, [4]University of Essex, *joint first authors

ed.donnellan@ucl.ac.uk, yumeng.wang.21@ucl.ac.uk

In naturalistic conversation, speakers often produce co-speech representational gestures, visually depicting properties of referents that they are talking about. These gestures commonly depict similar information as the accompanying speech, e.g., when talking about chopping a carrot, a speaker brings their hand down in a chopping action. In these instances, gestures may aid a speaker package conceptual information to facilitate language production: representing spatio-motoric features of a concept with the hands makes speech about that concept more accessible (Kita et al., 2017). On the other hand, interlocutors may predict upcoming speech from speakers' representational gestures, consistent with theories of multimodal integration of auditory and visual cues by interlocutors (e.g., Zhang et al., 2021). As such, co-speech representational gestures may facilitate language processing for both speaker and interlocutor.

Previous studies investigating the timing relationship demonstrate that, while relatively stable, there is still variability in the latencies between representational gestures and LAs (e.g., ter Bekke et al., 2020). This could be driven by two complementary sources: (i) differing demands on speaker production, e.g., word retrieval or (ii) audience effects, whereby a speaker modifies their communication sensitive to the requirements of their audience. In the current study we determine to what extent gesture-speech timing relationships vary depending on production demands or audience effects. To do so, we focused on co-speech representational gestures from the ECOLANG corpus (Vigliocco, et al., in prep) in which adult-adult (n=33) and adult-child dyads (n=36) engage in semi-naturalistic conversation about objects. We compared latencies between gesture stroke onsets (where the meaning of the gesture becomes clear) and LA onsets between gestures produced by adults to other adults (n=1928) and adults to their children aged 3-4 years (n=899). Crucially, to determine audience effects we compared between when the interlocutor was an adult or child. To investigate potential production demands, we considered the age of acquisition of the LA (AoA: indexing word retrieval difficulty) and whether the object was present or not (word retrieval may be easier when objects were physically present).

Our results suggest that audience effects play a large role in the timing of speaker production of representational gestures and their LAs. When interacting with children, gestures were produced later relative to the LA for later-acquired words, with later-acquired words produced concurrently or even before the gesture (Figure 1). This effect was more pronounced when objects being talked about were present (though the three-way interaction was non-significant). In contrast to previous work (e.g., Morrel-Samuels & Krauss, 1992), variability in timing does not seem to be as much due to demands on speaker production (e.g., the inherent difficulty of retrieving unfamiliar LAs). When interacting with other adults, object presence or AoA of the LA only had minor effects on the timing. To our knowledge, our work is the first demonstration that speakers flexibly alter representational gesture-speech timing relationships contingent on their interactional context.

Index Terms: audience effects; representational gestures; iconicity

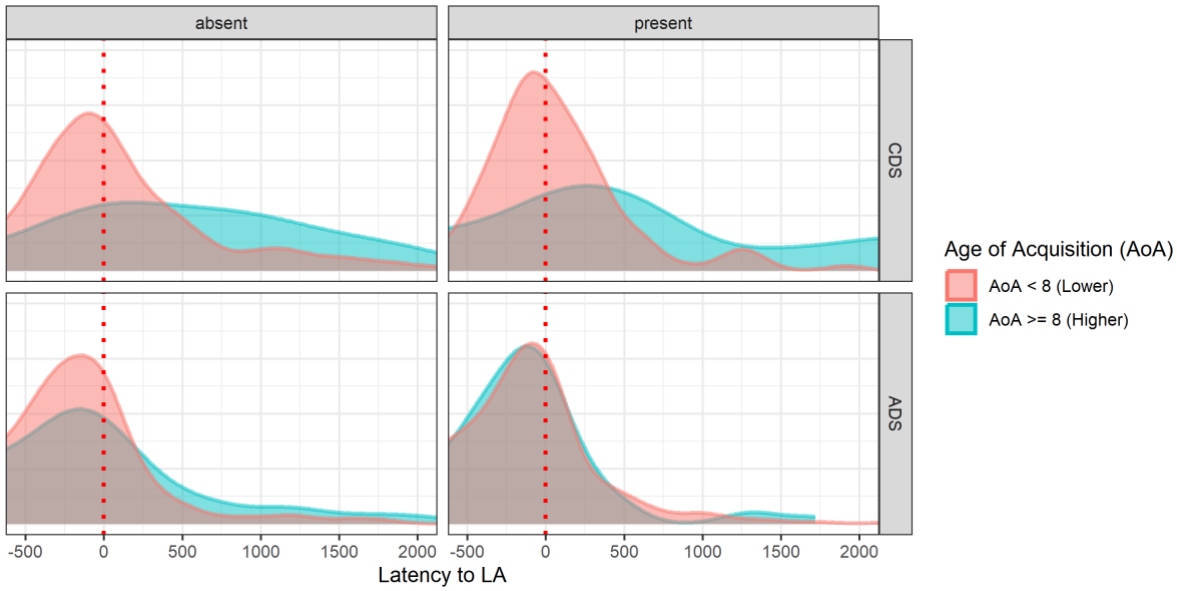

Figure 1: *Model predictions for Latency between LA and stroke*

References

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. Psychological Review, 124(3), 245–266. https://doi.org/10.1037/rev0000059

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(3), 615–622. https://doi.org/10.1037/0278-7393.18.3.615

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

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

Vigliocco, G., Gu, Y., Motamedi, Y., Grzyb, G., Brekelmans, M., Murgiano, M., Brieke, R., & Perniss, P. (unpublished). The ECOLANG corpus of dyadic interactions between caregivers and their 2–4 year-old child and between two adults.