Online processing of non-redundant iconic gestures: A visual world study

Isabella Fritz & Evelyn Milburn (Language Acquisition and
Language Processing Lab; Norwegian University of Science and Technology)

Iconic gestures and speech are integrated during comprehension in order to form a unified representation of a message (Özyürek, 2014). Information conveyed via iconic gestures is thought to add non-redundant meaning to an utterance (Alibali et al., 2009). However, previous online processing studies predominantly employed gesture-speech mismatch paradigms, where gestural information is also present in the speech channel. Thus, little is known about online processing of non-redundant information present in iconic gestures. We employed the visual world paradigm to investigate whether information exclusively conveyed via the gestural channel can be integrated with concurrent linguistic input.

36 participants viewed arrays of four images arranged in a grid (ex. orange, banana, disco ball, bench). Participants heard sentences like "/Someone will /peel/ a cheap/ fruit/" (slashes mark bins in later analyses). Each array was accompanied by one of three videos in the center of the screen. In the *Disambiguating* condition, the actor produced an iconic gesture that could depict two images in the array (ex. "round" gesture = orange and disco ball). However, the only image that fit a "round" gesture combined with the verb "peel" was the orange. The *Grooming* condition included a semantically empty hand movement. Gestures in both conditions started at the onset of the determiner (ex. *a* cheap) and always determined prior to the noun's onset. In the *No Gesture* condition, the actor sat at rest. If information conveyed via the non-redundant iconic gesture is fully integrated with linguistic input, we expect to see earlier and more looks to the verb+gesture-congruent object (orange) when gesture serves as a disambiguating cue. If non-redundant gestures are not integrated, there should be no difference in fixations between the two verb-congruent objects (orange/banana) regardless of gestural condition. Using Linear Mixed Effects Modeling, we compared fixation proportions in the four time bins marked by slashes in the example above, plus a bin that began at noun offset and ended 500 ms later (Figures 1a-c). We fitted separate models for each time bin including Condition as a three-level factor with the Disambiguating Condition as baseline (Comparisons: *Disambiguating – No gesture*; *Disambiguating – Grooming*). There were no significant effects in the first two bins. On the adjective, the Target (orange) advantage vs. the Verb Distractor (banana) was greater in the Disambiguating condition than the No Gesture condition ($\hat{\beta}$=.07; p<.05). However, the Target advantage vs. the Gesture Distractor (disco ball; $\hat{\beta}$=-.06; p<.05) and Unrelated Distractor (bench; $\hat{\beta}$=-.05; p<.05) was greater in the No Gesture condition than the Disambiguating condition (note that there were no differences in fixation proportions between the Target and the Verb Distractor in the Grooming and No Gesture conditions). On the noun, the Target advantage vs. the Verb Distractor was greater in the Disambiguating condition than in the Grooming condition ($\hat{\beta}$=.08; p<.05). In the post-noun bin, the Target advantage vs. the Verb Distractor was again greater in the Disambiguating condition than the No Gesture condition ($\hat{\beta}$=.15; p<.05).

Participants seemed to more easily converge on an appropriate image when gesture-speech integration wasn't necessary (results in the adjective bin). Thus, disambiguating gestures increased the processing load evident in participants not immediately converging on the target when presented with the gesture. These results are in line with previous ERP results showing that early ERP components (pre N400) are modulated similarly in complementary gesture conditions compared to gesture-speech mismatch conditions (Kelly et al., 2004). The degree of overlap between gestural and speech content thus influences how multimodal utterances are processed. Although some gestures facilitate communication (Holler & Levinson, 2019), non-redundancy in gestures leads to an increase in processing load. Despite the increase in processing load, our results from the last two bins indicate that non-redundant gestures were still integrated into their discourse context.

References

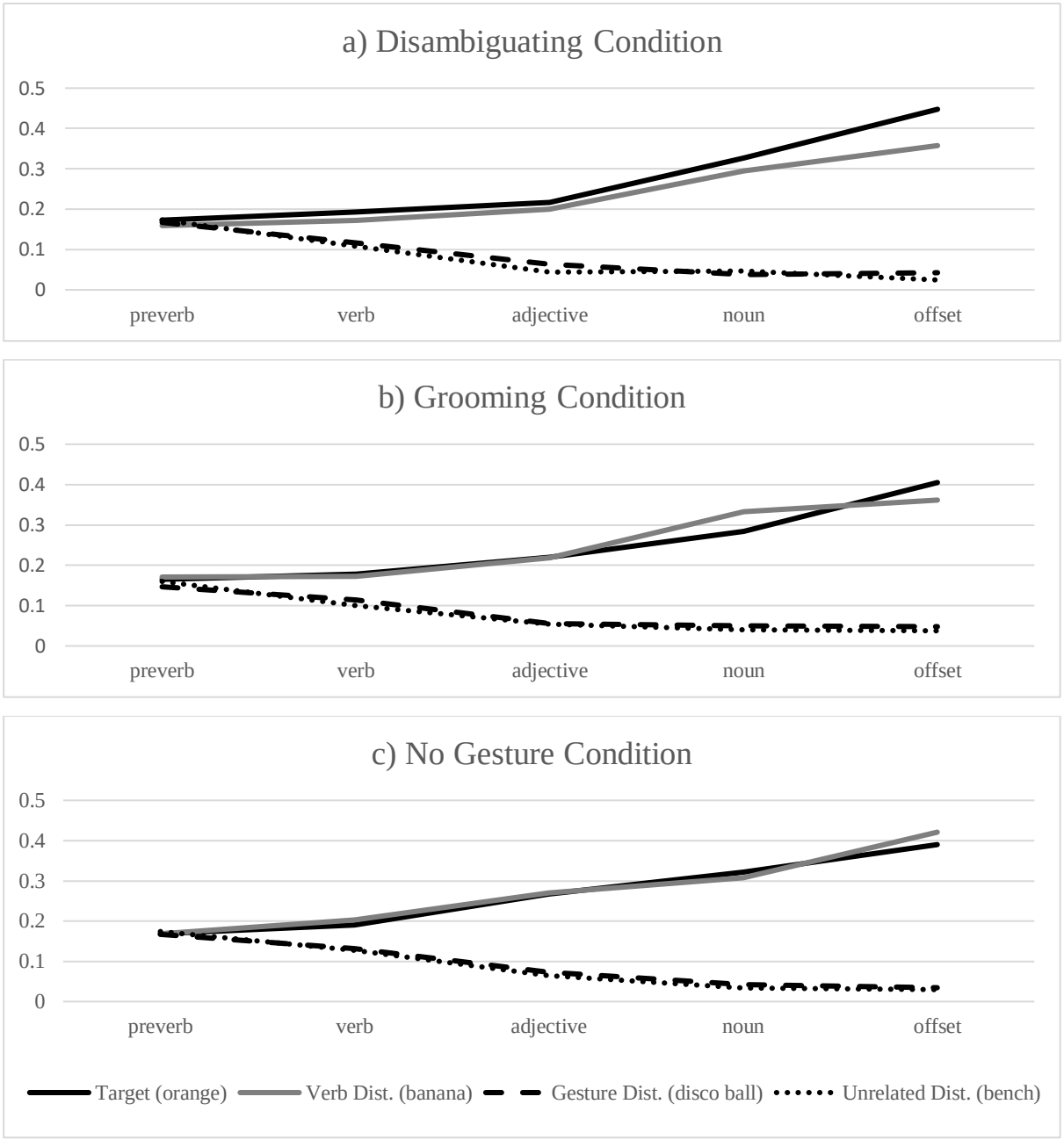
Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture–speech integration in narrative: Are children less redundant than adults?. *Gesture*, *9*(3), 290-311.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639-652.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, *89*(1), 253-260.

McNeill, D. (1992). Hand and Mind: What Gestures Reveal About Thought. Chicago: The University of Chicago Press.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philosophical *Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296.



Figures 1a-c: Fixation proportions in the three gestural conditions across time bins