Conceptual Integration and Multi-Modal Discourse Comprehension

Seana Coulson & Ying Choon Wu
University of California, San Diego

Iconic co-speech gestures are spontaneous body movements produced in coordination with speaking. For example, a speaker might trace an oval in the air as he says the word "platter". We suggest that speakers utilize conceptual integration, or blending, processes to combine linguistic information with visual-spatial and motoric information made available through gestures. Our model of multi-modal discourse makes a number of testable predictions. First, it suggests that gestural information enables speakers to formulate visually specific cognitive models of concrete discourse referents. Further, it suggests both visuo-spatial and sensori-motor working memory (WM) systems play an important role in speech-gesture integration.

In Experiment 1, 64 participants viewed short video clips of a speaker describing everyday objects and actions, followed by a photograph which was either related to the topic of the clip, or related to the topic of a different clip in the corpus. Half of the video clips were “congruous”, that is they were comprised of the original audio and video files; half were “incongruous”, constructed by pairing the audio file from one clip with the video file from another. Participants’ task was to attend to the videos, and, for each picture probe, press a button to indicate whether it was “Related” or “Unrelated” to the preceding video. After completion of the experiment, participants were subjected to a battery of tests to assess their verbal and visuo-spatial WM capacity.

Participants were faster to respond to probes following congruous than incongruous discourse primes, and to unrelated than related picture probes; the gesture congruity effect, however, was almost entirely attributable to responses to related picture probes. Consistent with our hypothesis that gestures convey visuo-spatial information about the discourse referents, these data suggest the difficulty of the picture probe task was affected by the congruity of the gestural information that accompanied the speech in the discourse primes. Moreover, consistent with our hypothesis that speech-gesture integration involves visuo-spatial WM, the size of the gesture congruity effect was significantly correlated with scores on a computerized version of the Corsi Block Task, an accepted index of visuo-spatial WM capacity.

Experiment 2 was intended to determine whether sensorimotor WM capacity was related to healthy adults’ ability to utilize semantic information in gestures. Accordingly, we assessed the sensorimotor WM capacity of 85 undergraduate participants in a customized version of the movement span task (adapted from [1]) as well as their performance on the picture probe task used in Experiment 1. Video primes were Congruent, Incongruent, or Neutral. Neutral videos were accompanied by an uninformative still image of the speaker. Responses to related pictures were significantly faster following Congruent videos than those following Neutral ones, and responses following Neutral videos were significantly faster than those following Incongruent videos. The size of this facilitation effect (the difference between Neutral and Congruent trials) was linearly related to participants’ movement span scores, suggesting sensorimotor WM systems are involved in speech-gesture integration processes.

These data suggest people recruit sensorimotor and visuospatial WM resources during speech-gesture integration.

References

[1] Smyth, M., & Pendleton, L. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology* (41A), 235-250.