Visual-Spatial and Verbal Working Memory Resources Modulate Multi-modal Co-reference

Stephani Foraker,
SUNY Buffalo State
forakesm@buffalostate.edu

Background

- Co-speech gestures – the hand and arm gestures people make while speaking – are tightly coordinated with the content of what they are saying (McNeill, 1992).
- Gestures can communicate information affecting the
  - meaning of nouns and verbs (Bernardis, Salillas & Caramelli, 2008)
  - position and size of objects (Holler, Shovelton, & Beattie, 2009)
  - comprehension of action verbs (Kelly, Ozyurek, & Maris, 2010)
- The location or hand shape of gestures that speakers spontaneously produce sometimes indicate co-reference between a pronoun and its referent (Foraker, 2010; So, Kita, & Goldin-Meadow, 2009).

Gestured information influences pronoun resolution, offline

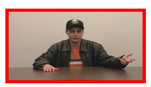
- Inhibitory effects: Gestures that contradict order of mention in a discourse can shift comprehender's interpretation of an ambiguous pronoun (Goodrich Smith & Hudson-Kam, 2012).
- Facilitative effects: Gestures consistently indicating an entity bias interpretation to that entity, whether first- or second-mention (Foraker & Delo, 2013 CUNY).

Gestured information guides pronoun resolution online

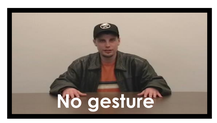
- The social cue of pointing to a referent present in the environment modulates pronoun resolution, shifting comprehender attention (Nappa & Arnold, 2014).
- Gestures in space, without referents present, also modulate pronoun resolution (Foraker, 2014 CUNY poster). When linking the pronoun with a referent representation (i.e., bonding, Garrod & Terras, 2000), we found that a gesture **consistent** with a referent **facilitated access**, but only for a **less accessible** referent (Foraker & McElree, 2007).
- suggests that working memory resources are involved in accessing a referent representation

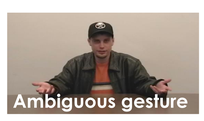
Design & Materials

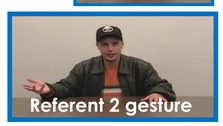
- First sentence introduced both referents; no gestures. Second sentence provided unique information about each referent, with an accompanying gesture as the name was uttered. Third sentence: **4 gesture conditions** – a gesture accompanies the pronoun



VIDEO: "Craig and Matt went on vacation.
Craig$_{[G1]}$ took a trip to Hawaii, while Matt$_{[G2]}$ took a trip to Florida.
He$_{[G1/NoG/AmbigG/G2]}$ thought the weather was great while on vacation."



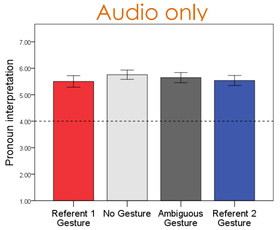




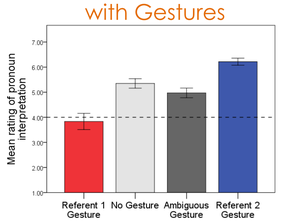
- First, 24 of 30 items were chosen from a written norm, where the pronoun was ambiguous with no preferred interpretation (scale below), and referent/pronoun gender was balanced.
- Videos were re-taped until rated naturalness of delivery and clarity of speech were equal in all conditions (4 naïve raters).
- Hand used was counterbalanced across order of mention; half deictic and half representational illustrator gestures; balanced across two speakers (1 M, 1 F)
- The extracted audio was first tested in the experimental design to ensure prosodic or other auditory information did not bias interpretation in our materials (32 participants, 24 items).
- Offline question: Who thought the weather was great while on vacation?
  - Craig for sure (1)…either one (4)…Matt for sure (7)

Offline judgments:

Audio only



with Gestures



- matching gestures biased interpretation
- No difference between deictic vs. iconic gestures
- 32 Ps, 24 items
- All comparisons significant by subjects and items, *ps* < .01

Funding from the Research Foundation, and the Office of Undergraduate Research, SUNY Buffalo State.
Thanks to members of the Cognition & Communication lab for assistance with developing the materials and data collection (Shukla, Nihlawi, Summers, Klipfel, Delo, Marshall, Hacker, Bellman).

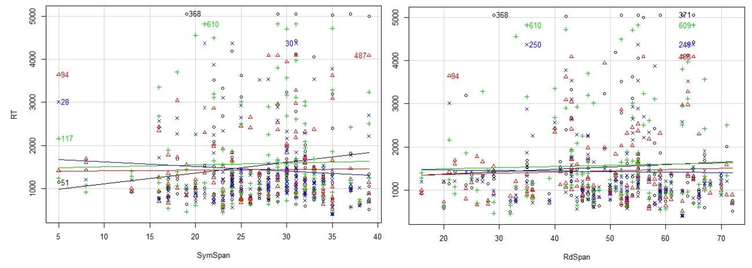
Predictions

- During comprehension, we predict that Verbal and Visual-Spatial Working Memory resources modulate multi-modal co-reference.
- Lower WM individuals benefit more from gesturing – during speech production
  - Gesture rate is higher for those with lower WM resources: Visual-spatial WM (Chu et al., 2014), Verbal WM (Gillespie et al., 2014)
  - Not allowing lower WM individuals to gesture reduces dual-task performance (Marstaller & Burianova, 2013)
- Higher WM supports resolving conflicting cues (King & Just, 1991; review, Engle, 2002), and integrating information more effectively (Ericsson & Kintsch, 1995).
- Here, we test to what degree higher WM may support integrating multi-modal information effectively.

Model fits: Linear mixed-effects regression, with predictors: Referent, Gesture, Reading Span *or* Symmetry Span score; Random slopes for subjects, items, and span score

Scatter Plot Legend:
No Gesture
Referent 1 Gesture
Referent 2 Gesture
Ambiguous Gesture

WM predicting Referent 1 recognition



- ✓ Both Visual-spatial and Verbal WM span predicted recognition time.
- ✓ Visual-spatial WM explained more variability than Verbal WM.
- ✓ Higher WM did not significantly affect integration of **matching** or **mismatching** gestures.
- ✓ When the gesture was **ambiguous**, those with higher visual-spatial WM were faster to recognize the referent. This may indicate better ability to retrieve the spatial location of the introducing gesture.
  - n.s. for verbal WM
- ✓ When **no gesture** was present, those with lower visual-spatial WM were faster to recognize the referent. This could indicate that mono-modal input (speech only) is easier for comprehension with constrained WM resources, and when retrieving a less accessible referent is needed.

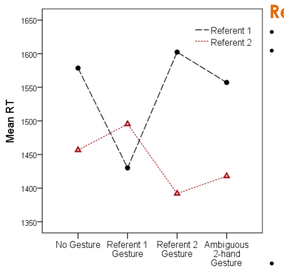
Online Co-reference: Referent Recognition Task

- Participants watched each video and decided if the name appearing above the video had been mentioned in that discourse or not.
- The name probe appeared at pronoun offset:
  - Referent 1 name (Craig), Referent 2 name (Matt)
  - a same gender foil (Brian), or an opposite gender foil (Susan)

Results

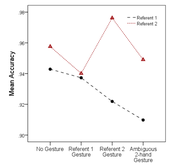
102 participants,
24 items, 16 lists

2 (referent) x 4 (gesture) ANOVA, wn-ss



Reaction Time Results

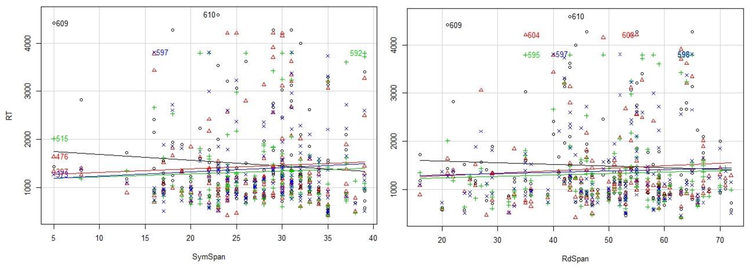
- Recency effect, $p = .002$
- Interaction of Referent x Gesture, $p = .067$
  - Referent 2 Name probes showed **no differences** between gesture types, *ps* > .173
  - For Referent 1 Name probes, reaction time was **faster** with a **consistent** Referent 1 gesture vs. an inconsistent Referent 2 gesture, $p = .046$.
- no main effect of gesture, $F < 1$



Accuracy Results

- Recency advantage, $p = .067$
- no main effect of gesture or interaction with referent, $Fs < 1$

WM predicting Referent 2 recognition



- ✓ WM explained less variability overall for the recent Referent 2.
- ✓ For those with lower WM, integration of **matching**, **mismatching**, and **ambiguous** gestures was facilitated compared to no gesture.
  - for visual-spatial WM more than verbal WM
  - no differences between matching, mismatching, and ambiguous gesture slopes
- ✓ Those with higher visual-spatial WM did not show differences between gesture conditions (nor for higher verbal WM).

Working Memory measures

(Redick et al., 2012)

- Verbal WM: Automated Reading Span score
- Visual-spatial WM: Automated Symmetry Span score

Discussion

- Gestures modulate online referent resolution, with interpretations shifted in either direction from baseline. Gestured content acts as one of several constraints during co-reference resolution.
- A consistent gesture facilitated access to the Referent 1 probe, but there was no effect of gesture for the uniformly faster Referent 2 probes. This suggests that gestured information can act as a retrieval cue for a referent representation that is less available in memory (Foraker & McElree, 2007).
- Visual-spatial WM resources modulated pronoun comprehension to a greater degree than Verbal WM.
- Both visual-spatial and verbal WM resources affected recognition of Referent 1. In particular, higher visual-spatial WM benefitted referent recognition when an ambiguous gesture was present, and lower visual-spatial WM individuals were best in the no gesture condition.
- WM resources did not affect Referent 2 recognition as much, although lower WM individuals did benefit from gestured information overall, while higher WM people showed no effect of gesture condition.