Beat gesture increases cognitive load during online contrastive reference resolution

Laura M. Morett (University of Alabama), Jennifer Roche (Kent State University), Scott H. Fraundorf (University of Pittsburgh), & James C. McPartland (Yale University)
lmorett@ua.edu

Introduction. In natural conversation, multiple cues to talker's intent are available, including both auditory and visual cues (e.g., gesture, lip movements) [1-2]. The availability of such multimodal cues raises the question of how they affect effort in comprehension: On one hand, processing cues beyond speech might increase cognitive load; on the other hand, additional cues to talker meaning might ease comprehension. Here, we use pupillometry to test these competing hypotheses in the context of viewing beat gesture during processing of spoken contrastive referring expressions. Prior work has shown that, relative to hearing presentational pitch accent, hearing contrastive pitch accent (CPA) used in conjunction with contrastive referring expressions decreases pupil size [3], consistent with evidence that pupil size increases are positively correlated with the difficulty of interpreting linguistic cues (e.g., connectives) during online sentence comprehension [4]. Here, we examine whether (a) viewing beat gesture exerts a similar effect during comprehension of spoken contrastive referring expressions and (b) whether any such effects were modulated by the felicity of beat gesture in context.

Methods. Participants (N = 40) performed a visual world task in which they heard pairs of sentences consisting of a context sentence followed by a critical sentence (1-4). Each sentence was accompanied by a centrally-presented video clip of a talker and an array of colored shapes (Fig. 1). Beat gestures consisted of downward flicks of one hand, and the side on which they occurred was counterbalanced to avoid side biases and object contingencies. In experimental trials, half of critical sentences referred to objects contrasting with the context object in color (1), and half referred to objects differing in both color and shape (2). Beat gesture and CPA were independently manipulated on the color adjective in these sentences. Experimental trials were randomly interleaved with filler trials, in which half of critical sentences contrasted with the context sentence in shape (3), and half differed in neither color nor shape (4). Because pupil size can be influenced by light levels, ambient light was kept consistent across trials, video luminance was standardized, and arrays were counterbalanced across trials to control for any differences in luminance of objects included in arrays. Pupil size was standardized because our eyetracking system (SR Research Eyelink) measures pupil size variation in units of eye-to-camera distance rather than in standard units (e.g., mm).

Results. Data were modeled with linear mixed-effect regression using the maximal random effect structures justified by the data for each model (R script: osf.io/ndh9r). A **manipulation check** using dwell time confirmed our analytic approach and ensured that participants attended to beat gesture: Participants spent more time looking at Targets in trials with beat gesture than in trials without beat gesture, (B = -0.03, t = -2.11, p = .04). The **main analysis** revealed that pupil size was larger in trials with beat gesture than without it (B = 0.02, t = 2.66, p = .008; see Table 1); however, pupil size did not differ by pitch accent or contrast type, and all interactions failed to reach significance. Additionally, pupil size decreased over time (B = -0.12, t = -2.24, p < .001) and reaction time was longer in trials with larger pupil size (B = 0.04, t = 4.13, p < .001).

Discussion. The results reveal that viewing beat gesture increases cognitive load during online resolution of spoken contrastive referring expressions. Notably, the effect of viewing beat gesture on cognitive load did not differ by contrast type (color vs. both), indicating that it reflects the intrinsic difficulty of integrating beat gesture with spoken contrastive referring expressions rather than the predictability or felicity of beat gesture in context. These findings suggest that, unlike auditory cues, visual cues may increase comprehenders' cognitive load via multimodality.

References. [1] Munhall, Jones, Callan, Kuratate, Vatikiotis-Bateson (2004). *Psych Sci.* [2] Sumby & Pollack (1954). *JASA.* [3] Zellin, Pannekamp, Toepel, & van der Meer (2011). *Int'l J. Psychophys.* [4] Demberg & Sayeed (2016). *PLoS One.*

Example sentences

1. Color-contrast (experimental): Click on the blue triangle. → Now click on the red triangle.
2. Both-contrast (experimental): Click on the blue square. → Now click on the red triangle.
3. Shape-contrast (filler): Click on the red square → Now click on the red triangle.
4. Neither-contrast (filler): Click on the red triangle → Now click on the red triangle again.

Table 1. Mean (standard deviation) pupil size by emphasis cue.

Emphasis	Pupil Size
Beat + CPA	274.95 (88.47)
Beat	273.55 (88.45)
CPA	271.48 (86.04)
No Emphasis	269.53 (88.21)

Figure 1. Screen layout for visual world task. (Object locations in array randomized for each trial and counterbalanced across participants.)