Pupillometry and Multimodal Processing of Beat Gesture and Pitch Accent: The Eye's Hole is Greater than the Sum of its Parts

Laura M. Morett[1] (lmorett@ua.edu)
[1]Department of Educational Studies in Psychology, Research Methodology, and Counseling
University of Alabama, Tuscaloosa, AL, 35401

Jennifer Roche[2] (jroche3@kent.edu)
[2]School of Health Sciences
Kent, OH 44242

Scott H. Fraundorf[3] (sfraundo@pitt.edu)
[3]Department of Psychology and Learning Research and Development Center
University of Pittsburgh, Pittsburgh, PA 15213

James C. McPartland[4] (james.mcpartland@yale.edu)
[4]Yale Child Study Center
New Haven, CT 06520

Abstract

This study investigated how beat gesture and pitch accent affect the cognitive load of listeners during language comprehension. Evidence from pupillometry and dwell time indicated that more cognitive resources were required to process the combination of these cues than their absence, and they suggest that beat gesture may have required more cognitive resources to process than pitch accent. Additionally, pupil size positively correlated with reaction time and decreased as the task progressed, demonstrating its usefulness as a measure of cognitive processing. These results indicate that viewing gesture in conjunction with speech may increase cognitive load during language processing, and that this increased load may result in enriched representations.

Keywords: beat gesture; pitch accent; cognitive load; language comprehension; visual world; pupillometry

Introduction

How does gesture that occurs with speech (co-speech gesture) affect the amount of mental effort (cognitive load) used during language processing? One prominent theory, advanced by Susan Goldin-Meadow and her colleagues (e.g., Goldin-Meadow, 2003), posits that gesture reduces cognitive load, freeing up cognitive resources for use in concurrent tasks such as speech production or problem solving. To date, most research examining gesture's effect on cognitive load has examined how gesture production affects cognitive task performance in speakers; only limited research has examined how gesture *viewing* affects cognition in listeners. One possibility that is consistent with Goldin-Meadow's hypothesis is that gesture viewing reduces cognitive load in comprehenders, similar to gesture production's effect on speakers, allowing comprehenders to devote more cognitive resources to concurrent tasks. We contrast this view with a novel alternative positing that combined gesture and speech require more cognitive resources to process than either alone. According to this alternative, viewing gesture in conjunction with speech should *increase* cognitive load because the combination requires more attentional resources to process than either alone; thus, fewer cognitive resources are available to devote to concurrent tasks. The current study evaluates these two possibilities by investigating how pupil size, an implicit measure of cognitive load, differs during comprehension of speech accompanied (or unaccompanied) by gesture.

Gesture's impact on cognitive load

Consistent with Goldin-Meadow's hypothesis, there is evidence that gesture production reduces speakers' cognitive load. For example, relative to not gesturing, gesturing while solving math problems improves memory for lists of words, letters, and grids presented prior to the problems (e.g., Wagner, Nussbaum, & Goldin-Meadow, 2004). Moreover, inhibition of gesture production impairs resolution of tip-of-the-tongue states (Frick-Horbury, 2002; Frick-Horbury & Guttentag, 1998), whereas gesture production facilitates it (Lucero, Zaharchuk, & Casasanto, 2014). Individuals with low working memory capacity produce more gestures during narrative recall than individuals with high working memory capacity (Gillespie, James, Federmeier, & Watson, 2014), and gesture production while solving math problems enhances recall of letters presented beforehand in individuals with low working memory capacity but not high working memory capacity (Marstaller & Burianová, 2013). These findings indicate that gesture production lightens cognitive load via physical action that enhances information processing, supporting Goldin-Meadow's hypothesis.

There is also evidence that gesture *viewing* facilitates cognition. For example, viewing teachers' gestures scaffolds learning of complex mathematical concepts such as equivalence and slope (Alibali & Nathan, 2007; Valenzeno, Alibali, & Klatzky, 2003), as well as L2 vocabulary

(Lazaraton, 2004). Moreover, work in which participants memorize lists of words from their native language (Igualada, Esteve-Gilbert, & Prieto, 2017; So, Chen-Hui, & Wei-Shan, 2012) or a novel second language (Allen, 1995; Bergmann & Macedonia, 2013; Kelly, McDevitt, & Esch, 2009; Kushch, Igualada, & Prieto, 2018; Macedonia, Müller, & Friederici, 2010; K.M. Mayer et al., 2015; Porter, 2012; Tellier, 2008) indicates that viewing gesture enhances memory for words (see Krönke et al., 2013 and Rowe, R.D. Silverman, & Mullin, 2013, for exceptions). Because all of this work examines how gesture viewing affects learning and long-term memory, however, it is less clear how it affects cognitive load in particular. Thus, additional investigation of gesture viewing's impact on real-time processing is needed to determine how it affects comprehenders' cognitive load during language processing.

Mechanistically, theories such as Dual Coding (Paivio, 1991) and Multimedia Learning (R.E. Mayer, 2002) posit that visual and verbal information are processed via complementary channels. Thus, these theories predict that processing conceptually-related information by viewing gesture and hearing speech simultaneously should not increase comprehenders' cognitive load beyond that required to process either cue alone. An alternative, described by the Split-Attention Effect (R.E. Mayer & Moreno, 1998), is that processing simultaneous redundant visual and verbal information requires greater attentional resources than processing information in either form alone. If this hypothesis is correct, comprehenders' cognitive load should be increased by processing gesture and speech simultaneously because this task requires division of attention between the two modalities. In each case, the presence of both visual and verbal information might result in enriched representations over the long term, but the real-time consequences for cognitive processing would differ.

Cues to emphasis

This study focuses on *beat gesture*, simple rhythmic body movements produced concurrently with speech. We examine beat gesture because, unlike other types of gesture, it has a singular purpose: to convey emphasis (McNeill, 1992; 2005). Beat gesture is closely related to pitch accent, a phonological construct that is realized acoustically as changes in the fundamental frequency, duration, and intensity of speech (Ladd, 1996): The timing of beat gesture is closely synchronized with the timing of pitch accent in production (Leonard & Cummins, 2011; Roustan & Dohen 2010), and words accompanied by beat gesture are more likely to be perceived as pitch accented than words unaccompanied by beat gesture (Krahmer & Swerts, 2007). Because beat gesture is closely related to pitch accent, it provides an opportunity to test whether providing similar information across modalities reduces or increases listeners' cognitive load during real-time language comprehension.

Unlike previous work examining gesture production's impact on cognitive load, which measured memory for lists of items unrelated to the main task, this study used pupil size (i.e., pupillometry) as its dependent variable. When other factors known to affect pupil size (e.g., luminance, attractiveness) are controlled, pupil size provides a direct, implicit, and real-time measure of attention and cognitive processing effort (Laeng, Sirois, & Gredeback, 2012). Thus, it is ideal for determining the cognitive load imposed by viewing gesture during language comprehension.

To our knowledge, this study is the first to use pupillometry to examine gesture's impact on cognitive load during language processing. Previous work has shown that listeners' pupil size decreases when processing sentences in which pitch accent is used contrastively (e.g., *Did Anne buy an umbrella? No,* Jenny *did*.), as opposed to sentences in which pitch accent is not used contrastively when it would be appropriate to do so (Zellin, Pannekamp, Toepel, & van der Meer, 2011). No such difference in pupil size was found for sentences with presentational pitch accent (e.g., *Who bought an umbrella?* Jenny *did*.), suggesting that contrastive pitch accent reduces comprehenders' cognitive load when used appropriately. This finding is consistent with work demonstrating that pupil size reflects the difficulty of interpreting other linguistic cues, such as connectives, during real-time sentence comprehension (Demberg & Sayeed, 2016). Moreover, it is consistent with work using other measures of processing, such as reaction time and grammaticality decisions, indicating that some linguistic structures and cues require more cognitive resources to process than others (Gibson, 1998; Lewis, Vashisth, & van Dyke, 2006; Wagers, Lau, & Phillips, 2009). In light of the close relationship between pitch accent and beat gesture, the findings of Zellin et al. (2011) suggest that pupillometry may be sensitive to gesture's impact on comprehenders' cognitive load during language processing, and thus should reflect the cognitive effort required to integrate gesture with speech.

To elucidate how beat gesture and pitch accent contribute to cognitive load individually and conjointly, we examined how pupil size differed when these cues were manipulated independently. To examine these processes in online reference resolution, we incorporated video stimuli into a visual world paradigm in which participants followed spoken instructions to interact with objects in a display. Similar tasks have been used successfully to examine how representational gesture is integrated with speech (L.B. Silverman, Bennetto, Campana, & Tanenhaus, 2010), how pitch accent affects sentence interpretation (Dahan, Tanenhaus, & Chambers, 2002; Ito & Speer, 2008; Watson, Tanenhaus, & Gunlogson, 2008), and how pitch accent affects cognitive load during language processing (Demberg & Sayeed, 2016). One possibility is that the combination of beat gesture and contrastive pitch accent recruits more cognitive resources than the absence of these cues. If this is the case, pupil size should be larger in trials with both beat gesture and contrastive pitch accent than in other trials. Additionally, we expected that pupil size would decrease over time and that increases in reaction time would predict increases in pupil size.

Method

Participants

Forty adults were recruited from the New Haven community to participate in this study and a related electrophysiological study in return for $25 compensation. All participants had normal hearing and normal or corrected-to-normal vision, and were not colorblind. Additionally, participants were screened for factors affecting pupil dilation (e.g., psychiatric and neurological disorders, drug consumption, medication).

Materials

A total of 656 sentences conveying simple instructions were audio recorded for this study (see 1a-2b for examples). 16 of these sentences were used for practice trials, and 640 were used for experimental trials. For both practice and experimental sets, half of the sentences were context sentences (1a), in which both the adjective and the noun had standard presentational pitch accents (PPA, H* in the ToBI system for intonational transcription of English; K. Silverman et al. 1992). The other half of the trials were continuation sentences, in which pitch accenting was manipulated on either the adjective or noun. Half of the continuation sentences were critical sentences, in which pitch accenting was manipulated on the adjective (2a), and half were filler sentences, in which pitch accenting was manipulated on the noun (2b). In critical sentences, the color adjective always differed from that of the preceding context sentence, and the noun was either the same or different, which was not pertinent to this analysis but was pertinent to another analysis. In half of the critical sentences, the adjective had contrastive pitch accenting (CPA; L+H* in ToBI), and in the other half, the adjective had PPA. Adjectives with CPA and PPA were spliced into identical carrier sentences (in which the original adjectives had PPA) to control the acoustic realization of the rest of the sentence. Similar to critical sentences, in half of filler sentences, the noun had CPA, and in the other half, the noun had PPA. Filler sentences were not spliced because eye gaze data were not examined during them. Table 1 summarizes the experimental design.

1a. *Context:* Click on the blue triangle.
2a. *Critical:* Now click on the *red* triangle/square.
2b. *Filler:* Now click on the blue *square/triangle*.

840 videos of a model producing the sentences described above were recorded to accompany them. 40 of these videos were used for practice trials, and 800 were used for experimental trials. 336 of these videos, which accompanied context sentences, did not contain beat gestures. In the other 504 videos, which accompanied continuation sentences, beat gesture was either present or absent in conjunction with either the adjective or noun. Two videos were recorded to accompany each critical sentence. In one of these videos, beat gesture was present in conjunction with the adjective; in the other, beat gesture was absent. Each filler sentence was always paired with the same video. Filler videos were constructed to maintain the association between beat gesture and pitch accent present in natural speech, such that beat gesture appeared with the noun in videos accompanying sentences with a contrastive pitch accent on the noun, but beat gesture was absent in videos accompanying fillers without a contrastive pitch accent.

Table 1: Experimental design (excluding practice trials).

Type	Contrast	Accent	Gesture	Trials
Context	N/A	PPA	None	320
Critical	Adj, both	CPA	Beat	40
Critical	Adj, both	PPA	Beat	40
Critical	Adj, both	CPA	None	40
Critical	Adj, both	PPA	None	40
Filler	Noun	CPA	Beat	80
Filler	Neither	PPA	None	80

All videos were recorded separately from audio and were aligned temporally with audio in post-production. A total of 64 colored shapes (8 colors x 8 shapes) were created for use in arrays accompanying audio and video stimuli. During the experiment, videos were presented centrally with a circular mask, and shapes were positioned equidistant from the center of the screen in a square configuration (see Figure 1).

Procedure

Before beginning the experimental task, participants were seated 55-56 cm from the screen (35° 55’ 0.32” visual angle). Gaze was calibrated to within 0.5° of visual angle using 13 points of reference. Drift checks and recalibration were performed between experimental trial blocks.

At the beginning of the experimental task, participants were told that it tested their ability to follow instructions. Participants were instructed to respond to all instructions issued in the paradigm by clicking on the appropriate shape. In the event they clicked on the wrong shape, participants were instructed to click on the correct shape to proceed. All responses for critical experimental trials were correct.

To become familiar with the task, participants first completed a practice phase of 8 trials. Participants then proceeded to the experimental phase, which consisted of four blocks of 40 trials. In both phases, critical and filler

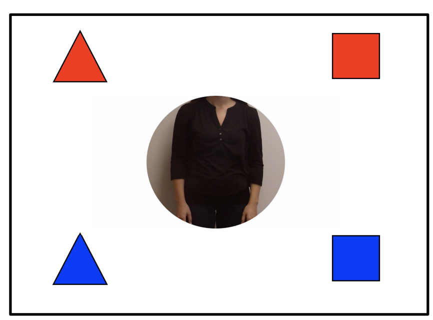

Figure 1: Schematic of screen configuration.

trials were randomly intermixed. In each trial, an array of shapes appeared and a video began playing, and the context sentence was presented aurally after a 200 ms delay. This configuration ensured that the apex of the beat gesture occurred 200 ms prior to the onset of the corresponding word, which is consistent with the timing of gesture production relative to speech (Morrel-Samuels & Krauss, 1992) and perceptual biases for gesture relative to speech (Leonard & Cummins, 2011). Following a correct response, the video was replaced by a gray circular placeholder for 1000 ms while the array of shapes remained on screen. Subsequently, the sequence repeated with the continuation sentence and corresponding video. Following a correct response, the trial ended and, after a blank screen was displayed for 1000 ms, a new trial began. Gaze data was collected remotely from the right eye at a 500 Hz sampling rate using an EyeLink 1000 eyetracker. Additionally, response latency and accuracy data were collected.

Measures

Dwell Time has been used as a measure of cognitive processing effort, with longer dwell times indicating greater effort (e.g., Ehrlich & Rayner, 1981; Rayner & Duffy, 1986). This measure refers to the average length of time the participant looks at particular interest areas—in this case, the gesture video and the target shape—as the participant hears the continuation sentence unfolding over time. This measure was used as a manipulation check, as it was expected that participants would look longer at the video because it was centrally-presented and dynamic.

Pupil Size, or pupillometry, is a measure that reflects non-volitional cognitive processing effort and attention (Laeng, Sirois, & Gredeback, 2012). Because pupil size can be influenced by light levels, we kept the ambient light consistent across trials, standardized video luminance (0.34-0.76 IRE), and counterbalanced shape arrays across trials to control for any differences in luminance. Pupil size was standardized because Eyelink's software measures pupil size variation in units derived from eye-to-camera distance rather than in standard units of measurement (e.g., mm).[1]

Results

We used linear mixed effect models with the lme4 package in R to evaluate changes in dwell time (model 1: manipulation check), and pupillometry (model 2; find R scripts here: osf.io/fy6wp). Each model implemented the maximal random effect structure permitting model convergence, with participant included as the random effect.

Model 1: Dwell Time Manipulation Check evaluated differences in dwell time to two areas of interest: video and target shape location. This manipulation check was conducted to ensure that participants were attending to the gestures, as reflected in greater dwell times on the video. This model evaluated dwell time (DV) as a function of emphasis (Beat + CPA, Beat, CPA, No Emphasis), interest area (Video or Target Shape), and their interaction as fixed effects (marginal R^2 = .22, condition R^2 = .54). A significant main effect of interest area indicated participants spent more time looking at Videos than Target Shapes, (B = .91, t = 5.51, p < .001). There was also a marginal difference between Beat + CPA and No Emphasis; participants spent marginally more time looking at Videos as well as Target Shapes when both beat gesture and CPA were present than when both of these cues were absent, (B = -.06, t = -1.68, p = 0.09). This suggests that more cognitive resources may be recruited to process the combination of beat gesture and CPA than neither cue to emphasis.[2] To further test this hypothesis, the next model evaluated changes in pupil size by emphasis, time, and reaction time.

Model 2: Changes in Pupil Size were evaluated by emphasis, time, and reaction time[3]. In this model, emphasis, time (trial), and reaction time were set as fixed effects (marginal R^2 = .02; conditional R^2 = .84). The results indicated that pupil size was larger in trials with beat gesture and CPA relative to trials with neither of these cues (B = 0.02, t = -2.40, p = .02), and marginally larger relative to trials with CPA but no beat gesture (B = 0.01, t = 1.85, p = .06; see Table 2 for descriptive statistics). Additionally, pupil size decreased over time (B = -0.04, t = -15.67, p < .001), and reaction time was longer in trials with larger pupil size (B = 0.01, t = 3.41, p < .001).

Discussion

Consistent with our predictions, both viewing beat gesture and hearing pitch accent increased listeners' cognitive load, as evidenced by marginally longer dwell time and larger pupil size. Of the two cues, beat gesture appeared to be the larger contributor, given that pupil size was marginally larger for the combination of beat gesture and CPA relative

Table 2: Mean (standard deviation) pupil size by emphasis. Difference from Beat + CPA: *$p \geq .05$; †$p \geq .09$

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)*

[1] As reported by Hayes and Petrov (2017), Eyelink Dataviewer software calculates pupil size to "form a ratio scale that is layout-dependent of proportionality to millimeters" (p. 5).

[2] Because this effect was found for looks to the target shape as well as the video and was not found for beat gesture alone, it is unlikely that it is driven solely by the motion associated with beat gesture.

[3] Reaction time was evaluated because it is another measure sometimes used to measure cognitive load (Barrouillet et al., 2007). Additionally, reaction time was independently evaluated and produced no significant effects of emphasis type.

to CPA alone, and that there was no difference in pupil size between the combination of beat gesture and CPA and beat gesture alone.

Considered together, these results suggest that emphasis conveyed via both beat gesture and contrastive pitch accent requires more cognitive resources to process than emphasis conveyed via CPA alone. This finding is consistent with the Split-Attention Effect (Mayer & Moreno, 1998), suggesting that simultaneous occurrence of cues to emphasis in the visual and verbal modalities may initially tax listeners' cognitive resources, increasing cognitive load relative to either type of cue alone. Importantly, previous work demonstrates superior memory for verbal information accompanied by both beat gesture and CPA relative to information accompanied only by contrastive pitch accent when both cues are present in some cases, but not others (Kushch & Prieto, 2016; Morett & Fraundorf, 2016). Thus, the additional cognitive resources necessary to process both beat gesture and CPA may result in enriched memory traces, consistent with the predictions of Dual Coding and Multimedia Learning Theories and "Desirable Difficulty" theories of memory (e.g., Schmidt & Bjork, 1992).

Notably, this study was the first to use pupillometry to examine how viewing gesture affects listeners' cognitive load during real-time language comprehension. The results indicate that changes in pupil size provide a reliable measure of processing effort, given that pupil size and reaction time were positively related (i.e., longer reaction time was associated with increased pupil size). Furthermore, as the task progressed, pupil size decreased, suggesting that cognitive processing increased in efficiency with repetition. Considered in conjunction with the findings of other work using pupillometry to examine variation in cognitive load during language processing (Demberg & Sayeed, 2016; Zellin et al., 2011), this work demonstrates that pupillometry provides insight into the use of cognitive resources during real-time gesture-speech integration.

Acknowledgements

This research was funded by a Hilibrand Postdoctoral Fellowship to L.M.M and NIMH R01 MH107426 to J.C.M. The authors thank Talena Day, Kathryn McNaughton, and Zachary Williams for their help with data collection.

References

Alibali, M. W., & Nathan, M. J. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. *Video Research in the Learning Sciences*, 349-365.

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal, 79,* 521–529.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570.

Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: iconic gestures help to improve learners' memory performance. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *International Workshop on Intelligent Virtual Agents: IVA 2013* (pp. 139-148). Berlin: Springer.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292-314.

Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLOS One, 11,* e0146194.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641-655.

Roustan, B., & Dohen, M. (2010). Co-production of contrastive prosodic focus and manual gestures: temporal coordination and effects on the acoustic and articulatory correlates of focus. In *Speech Prosody 2010*. Chicago: International Speech Communication Association.

Frick-Horbury, D. (2002). The use of hand gestures as self-generated cues for recall of verbally associated targets. *The American Journal of Psychology, 115,* 1–20.

Frick-Horbury, D., & Guttentag, R. E. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology, 111,* 43–62.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68,* 1-76.

Gillespie, M., James, A. N., Federmeier, K. D., & Watson, D. G. (2014). Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition, 132,* 174-180.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*, 510-527.

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *Journal of Experimental Child Psychology*, *156*, 99–112.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541-573.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*, *57*, 396–414.

Krönke, K. M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex, 49,* 2553-2568.

Kushch, O., & Prieto, P. (2016). The effects of pitch accentuation and beat gestures on information recall in

contrastive discourse. In *Speech Prosody 2016* (pp. 922–925). Boston: International Speech Communication Association.
Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience*, 1–13.
Ladd, D. R. (1996). *Intonational phonology*. New York: Cambridge University Press.
Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious?. *Perspectives on Psychological Science*, *7*, 18-27.
Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning*, *54*, 79-117.
Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences, 10,* 447-454.
Loewenfeld, I. (1993). *The pupil*. Detroit: Wayne State University Press.
Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26,* 1457–1471.
Lucero, C., Zaharchuk, H., & Casasanto, D. (2014). Beat gestures facilitate speech production. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual meeting of the Cognitive Science Society*. Red Hook, NY: Curran Associates.
Macedonia, M., Müller, K., & Friederici, A. D. (2010). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping,* 32, 982–998.
Marstaller, L., & Burianová, H. (2013). Individual differences in the gesture effect on working memory. *Psychonomic Bulletin & Review, 20,* 496-500.
Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology, 25,* 530-535.
Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41,* 85-139.
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*, 312.
McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
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,* 615.
Morett, L. M. & Fraundorf, S. H. (2016). The impact of beat gesture and pitch accenting on memory for spoken discourse. Presented at the 57th annual meeting of the Psychonomic Society, Boston, MA.
Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.
Porter, A. (2012). A helping hand with language learning: teaching French vocabulary with gesture. *The Language Learning Journal, 44,* 236-256.
Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191-201.
Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology, 38,* 109-117.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., et al. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the third international conference on spoken language processing* (pp. 867–870).
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207-218.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., ... & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *ICSLP-1992,* 867-870.
Silverman, L. B., Bennetto, L., Campana, E., & Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition, 115,* 380-393.
So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?. *Language and Cognitive Processes, 27,* 665-681.
Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8,* 219-235.
Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*, 187-204.
Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61,* 206-237.
Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial?. *Journal of Memory and Language, 50,* 395-407.
Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+ H. *Cognitive Science, 32,* 1232-1244.
Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology, 81,* 133-141.