Eye See What You're Saying: Beat Gesture Facilitates Online Resolution of Contrastive Referring Expressions in Spoken Discourse

Laura M. Morett[1] (lmorett@ua.edu), **Scott H. Fraundorf**[2] (sfraundo@pitt.edu),
James C. McPartland[3] (james.mcpartland@yale.edu)
[1]Department of Educational Studies, University of Alabama, Tuscaloosa, AL 35401
[2]Department of Psychology and LRDC, University of Pittsburgh, Pittsburgh, PA 15260
[3]Yale Child Study Center, New Haven, CT 06520

Abstract

This study investigated how beat gesture and contrastive pitch accenting affect online contrastive reference resolution during spoken discourse comprehension. Evidence from gaze fixations indicated that beat gesture encouraged fixations to target referents of contrastive referring expressions and that contrastive accenting encouraged fixations to competitor referents of non-contrastive referring expressions. Notably, beat gesture and contrastive accenting acted independently, indicating that their effects are additive rather than interactive. Moreover, neither beat gesture nor contrastive accenting affected an observed tendency to anticipate contrastive referring expressions. Together, these results provide the first evidence that beat gesture, like contrastive accenting, is interpreted as a cue to contrast during online reference resolution in spoken discourse comprehension.

Keywords: beat gesture; pitch accent; reference resolution; discourse processing; visual world; eye tracking

Introduction

Successful discourse comprehension entails establishing relations between entities. One such relation is *contrast,* which refers to a contradiction between two themes (Myhill & Xing, 1996). An example can be seen in the distinction between referents in the following discourse: *The report isn't due on Tuesday; it's due on Thursday*. Although contrast can be discerned semantically, cues conveying prominence can be used to highlight it, strengthening the propositional representations of both the speaker and the listener. Two such cues are pitch accent—alterations in in speech fundamental frequency (f0), duration, and intensity (Ladd, 1996)—and beat gesture—simple rhythmic gesture (McNeill, 1992; 2005). Although processing of these cues has been studied in offline discourse comprehension (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018), it is currently unclear how it affects *online* discourse comprehension. The current study uses eyetracking to examine how independently manipulating pitch accent and beat gesture affects online contrast interpretation in spoken discourse. In doing so, it provides insight into the individual and combined contributions of these cues to prediction and resolution of contrast in particular, as well as representation and processing of inter-entity relations more generally, in spoken discourse.

Cues to contrast

Two of the most prominent types of pitch accenting in English discourse are presentational pitch accenting (PPA), which is used to convey new, non-contrastive information, and contrastive pitch accenting (CPA), which is used to convey information contrasting with other mentioned information. These two pitch accents differ acoustically; PPA (H* in the ToBI framework) consists of a high pitch target and f0 high in the talker's range, whereas CPA (L+H* in the ToBI framework) consists of an initial low pitch followed by a sharp rise to a high target on the accented syllable (Beckman & Elam, 1997; K. Silverman et al., 1992). Previous work demonstrates that listeners are sensitive to the distinction between PPA and CPA, and this is reflected in both memory for discourse and real-time discourse comprehension. Referents with CPA are remembered better than referents with PPA, particularly when a salient contrasting item must be rejected (e.g., remembering *Scottish* rather than *British;* Fraundorf et al., 2010; 2012; Lee & Fraundorf, 2016; Lee & Snedeker, 2016; Sanford, Sanford, Molle, & Emmott, 2006). Moreover, CPA facilitates rejection of items contrasting with contrastively-accented referents (e.g., *dish* given *antenna*), but not objects with non-contrastive relations to those referents (e.g., *television* given *antenna*; Braun & Tagliapietra, 2010). Lastly, in eyetracking studies, CPA encourages anticipatory looks to objects contrasting with previously-mentioned referents (e.g., after hearing "red scissors," to purple scissors upon hearing "PURPLE"), even when the referent is subsequently revealed to be non-contrastive (e.g., *book*; Ito, Jincho, Minai, Yamane, & Mazuka, 2012; Ito & Speer, 2008; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Watson, Tanenhaus, & Gunlogson, 2008; Weber, Braun, & Crocker, 2006).

Like pitch accenting, beat gesture is used to emphasize important information in spoken discourse, such that it serves as a "yellow gestural highlighter" (McNeill, 2006). Indeed, both alone and in combination with pitch accenting, beat gesture enhances memory for information conveyed via discourse (Austin & Sweller, 2014; Igualada, Esteve-Gilbert, & Prieto, 2017; Morett, 2014; Vilà-Giménez, Igualada, & Prieto, in press). Moreover, some work indicates that beat gesture enhances memory for contrastive information in particular, especially when it occurs in conjunction with CPA (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018). These findings suggest that beat gesture strengthens memory traces for information in spoken discourse by increasing its salience visually. In addition to their similarity in function, beat gesture and pitch accenting are closely related in timing. Indeed, beat gesture and pitch accenting are temporally aligned on both the sentential and syllabic levels (Esteve-Gilbert & Prieto, 2013; Leonard & Cummins, 2011), suggesting that the temporal relationship between these two cues to prominence is based on prosody.

Considered as a whole, these findings demonstrate that beat gesture and pitch accenting are closely related in timing and meaning. This suggests that beat gesture—as another cue to prominence—might facilitate online processing of contrast in spoken discourse, as CPA does, Further, the functional similarity of beat gesture and CPA highlights the need to investigate their effects on online contrast processing not only independently but also *conjointly*.

Cue integration

Despite the similar function and close relationship of pitch accenting and beat gesture, relatively little research has examined how the presence—and absence—of these cues in relation to one another affects interpretation of contrast in spoken discourse. In a focus production task in which participants produced beat gesture and/or pitch accenting on one or both referents of a sentence (*Amanda goes to Malta*), referents produced with beat gesture alone had higher vowel formants and were more likely to be perceived as pitch accented than referents unaccompanied by beat gesture (Krahmer & Swerts, 2007). However, in a similar task that involved producing beat gesture in conjunction with contrastive corrections after hearing sentences (*Baba holds the baby? → Mumu holds the baby*), beat gesture production did not affect the articulatory or acoustic correlates of CPA (Roustan & Dohen, 2010). Moreover, in a focus comprehension task in which pitch accenting and beat gesture were independently manipulated in conjunction with the patients of transitive sentences (e.g., *Yesterday, Anna brought fresh lilies to the room*), pitch accenting elicited a larger N400 response when beat gesture was absent than when beat gesture was present, indicating greater inconsistency with predictions or difficulty of semantic integration in the former case (Wang & Chu, 2013). Taken together, these findings indicate that the co-occurrence patterns of pitch accenting and beat gesture affect their interpretation as cues to contrast in spoken discourse.

The influence of beat gesture on interpretation of pitch accenting is also evident in work indicating that information conveyed via spoken discourse accompanied by both beat gesture and pitch accenting is remembered better than the same information accompanied by pitch accenting alone. This result has been observed for memory of focal, non-contrastive information (Igualada, Esteve-Gibert, & Prieto, 2017; Kushch, Igualada, & Prieto, 2018; Morett, 2014; Vilà-Giménez, Igualada, & Prieto, in press) as well as contrastive information (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018; Morett & Fraundorf, under review). With respect to memory for contrastive information, the authors' previous work indicates that, when both cues are manipulated independently in a within-subjects design, contrastive information with CPA is remembered better than contrastive information with PPA when beat gesture is present, but not when beat gesture is absent. When beat gesture is *never* present, however, contrastive information with CPA is remembered better than contrastive information with PPA, consistent with the findings of previous work demonstrating the same effect using similar paradigms presented only in the auditory modality (Fraundorf et al., 2010; 2012; Lee & Fraundorf, 2016; Lee & Snedeker, 2016; Sanford, Sanford, Molle, & Emmott, 2006). Considered as a whole, these findings suggest that beat gesture and CPA influence one another in offline discourse comprehension and memory. However, it is less clear whether and how these cues interact in online discourse processing.

To elucidate how beat gesture and CPA affect contrastive reference resolution in online spoken discourse, we examined differences in fixations to referents accompanied by beat gesture and/or CPA. To do so, we used a modified version of the visual world paradigm that included video. The visual world paradigm has been used successfully to examine how CPA affects online reference resolution (Ito & Speer, 2008; Kurumada et al., 2014; Watson, Tanenhaus, & Gunlogson, 2008), as well as how representational gesture is integrated with speech online (L. B. Silverman, Bennetto, Campana, & Tanenhaus, 2010). Based on these studies and the related work discussed above, we predicted that beat gesture and CPA would affect online reference resolution. Specifically, we predicted that, when referents contrasted only in color (e.g., *blue triangle* and *red triangle*), the presence of beat gesture alongside the color word would facilitate reference resolution, particularly in conjunction with CPA. By comparison, we predicted that when referents differed in both color and shape (e.g., *blue square* and *red triangle*), the presence of beat gesture alongside the color word would misleadingly suggest a color contrast and hinder reference resolution, particularly in conjunction with CPA.

Methods

Participants

Forty adult native English speakers (age range: 18-35 years; 29 females, 11 males) were recruited to participate in this study on a paid basis. All participants had normal hearing and normal or corrected-to-normal vision and were not colorblind. Additionally, participants were screened for factors affecting eye movements (e.g., psychiatric and neurological disorders, recreational drug use).

Materials

A total of 672 referring expressions conveying simple instructions were audio recorded (see 1a-2b for examples; 32 practice, 640 experimental). In both practice and experimental trials, half of referring expressions provided context, with standard PPA on both color and shape words. The other half of referring expressions provided continuation, consisting of half critical and half filler trials. In critical trials, the color word always differed from that of the preceding context referring expression, and the shape word was either the same (color-contrast; 1a) or different (both-contrast; 1b). In both types of critical trials, pitch accenting was manipulated by splicing color words with CPA or PPA into identical carrier sentences (in which original color and shape words had PPA) to control acoustic realization of the rest of the referring expression. Filler trials were created to represent the other possibilities, in which the color word was the same as that of the preceding context referring expression and the shape word either differed (shape-contrast; 2c) or was the same (no-contrast; 2d). In these trials, pitch accent was always felicitous, such that shape words in shape-contrast referring expressions always had CPA and shape words in no-contrast referring expressions always had PPA. Sentences in filler trials were recorded wholesale and were not spliced.

1a. *Color-contrast:* Click on the blue triangle → red triangle.
1b. *Both-contrast:* Click on the blue square → red triangle.
2a. *Shape-contrast:* Click on the red square → red triangle.
2b. *No-contrast:* Click on the red triangle → red triangle.

840 videos of a talker producing the sentences described above were recorded to accompany audio recordings. 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 referring expressions, beat gesture was either present or absent alongside the color word (for critical trials) or shape word (for filler trials). Two videos were recorded to accompany each critical referring expression. In one of these videos, beat gesture was present alongside color words; in the other, beat gesture was absent. Videos recorded to accompany filler trials maintained the association between beat gesture and CPA present in natural speech; in videos accompanying shape-contrast referring expressions), beat gesture occurred alongside CPA-accented shape words, whereas beat gesture was absent from videos accompanying PPA-accented no-contrast referring expressions. A within-participants design was used such that each participant received all combinations of contrast type, beat gesture, and pitch accenting; however, combinations for individual trials were counterbalanced across participants (see Table 1 for experimental design summary). All videos were recorded separately from audio and were aligned temporally with it in post-production. Because beat gestures were produced with one hand, consisting of a single downward stroke with the palm oriented upward, horizontally-flipped duplicates were created for all videos in post-production.

A total of 64 objects (8 colors x 8 shapes) were created for inclusion in arrays accompanying audio and video stimuli. Videos were presented centrally with a circular mask, with objects positioned equidistantly (see Fig. 1). Locations in which objects appeared were counterbalanced to control for contingencies between them and beat gesture orientation.

Table 1: Experimental design (excluding practice trials).

Type	Contrast	Accent	Gesture	Trials
Critical	Color	CPA	Beat	20
Critical	Color	No CPA	Beat	20
Critical	Color	CPA	None	20
Critical	Color	No CPA	None	20
Critical	Both	CPA	Beat	20
Critical	Both	No CPA	Beat	20
Critical	Both	CPA	No Beat	20
Critical	Both	No CPA	No Beat	20
Filler	Shape	CPA	Beat	40
Filler	None	No CPA	None	40

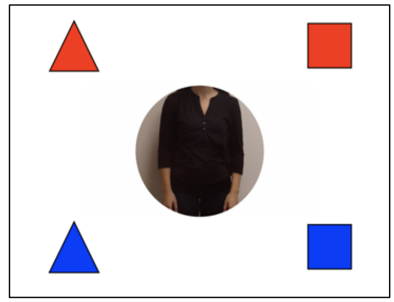

Figure 1: Schematic of screen configuration.

Procedure

Fixation data was collected remotely from the right eye at a 500 Hz sampling rate using an EyeLink 1000 eyetracker. 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 recalibrations were performed between experimental trial blocks.

At the beginning of the experimental task, participants were told that its objective was to test their ability to follow instructions. Participants were told to respond to all instructions issued in the paradigm by clicking on the appropriate object. The experiment was programmed such that participants who clicked on the wrong object were instructed to click on the correct object to proceed. However, all responses to critical referring expressions were correct. This was not surprising given that the task was simple and straightforward, as is characteristic of visual world tasks (Huettig, Rommers, & Meyer, 2011; Salverda, Brown, & Tanenhaus, 2011), and our intent was to assess the online processing of correctly-understood referring expressions.

To become familiar with the experimental task, participants first completed a practice phase consisting of 8 trials. Participants then proceeded to the experimental phase, which consisted of four blocks of 40 trials each. In both phases, critical and filler trials were randomly interleaved. In each trial, an array of objects appeared and a video began playing, and the context referring expression 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 in natural discourse (Morrel-Samuels & Krauss, 1992) as well as perceptual biases for the timing of beat gesture relative to speech (Leonard & Cummins, 2011). Following a correct response, the video disappeared and was replaced by a gray circular placeholder for 1000 ms while the object array remained on screen. Subsequently, the sequence repeated with the continuation referring expression and corresponding video. Following a correct response, the trial ended and, after a blank screen was displayed for 1000 ms, a new trial began.

Results

We examined fixations during two periods of the critical referring expression: *color word* (color word onset to shape word onset) and *shape word* (shape word onset to response onset). To account for saccade planning, each period was shifted ahead by 200 ms. Two interest areas relevant to the main research question were defined: the *target object* referred to by the critical referring expression, and the *competitor object* that is temporarily consistent with the unfolding linguistic input. In color-contrast trials, in which the target object contrasted with the referent of the context referring expression only in color, the competitor object differed in both color and shape; in both-contrast trials, in which the target object differed from the referent of the context referring expression in both color and shape, the competitor object contrasted only in color. In addition, the *video* interest area was defined to confirm that participants were watching the video. Participants fixated the video more than target and competitor objects combined during both the color word (77.65% of fixations) and shape word (52.19% of fixations) interest periods.

To account for non-independence of samples, we summed fixations within each interest period in each trial and took the empirical logit (Barr, 2008). Because we were interested in how beat gesture and CPA facilitated and hindered reference resolution, empirical logit values were computed for fixations to target and competitor objects separately. These values were then entered into linear mixed effects models, which were fit using the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2015) and evaluated via null hypothesis statistical testing using the *lmerTest* R package (Kuznetsova, Brockhoff, & Christensen, 2017). Each model implemented the maximal random effect structure permitting convergence, with beat gesture, CPA, contrast, and their interactions as fixed effects and participant and trial as random effects. To account for any effects of spatial orientation of the gesture, we also included gesture orientation and target object side as control variables.

Color Word Interest Period

For target fixations, we observed a main effect of contrast, indicating a higher likelihood of fixating target objects during color-contrast (M = -0.11, SD = 0.55) than both-difference critical referring expressions (M = -0.17, SD = 0.55; t = -3.57, p < .001); however, no interactions between contrast and either accent or gesture were observed. For competitor fixations, no main effect of contrast was observed (color: M = -0.16, SD = 0.54; both: M = -0.09, SD = 0.57; t = 1.62, p = .11). Moreover, there were no main effects or interactions of gesture orientation and target side for target or competitor fixations. Thus, although these results suggest a baseline bias towards contrastive interpretation of critical referring expressions, neither beat gesture nor CPA enhanced resolution of these expressions prior to disambiguation.

Shape Word Interest Period

We observed significant two-way interactions between orientation and target side for both target and competitor fixations (target: B = 0.48, SE = 0.08, t = 6.32, p < .001; competitor: B = -0.10, SE = 0.04, t = -2.58, p = .01), indicating a baseline tendency to fixate target objects and not to fixate competitor objects appearing on the side of the array congruent with the orientation of accompanying beat gesture.

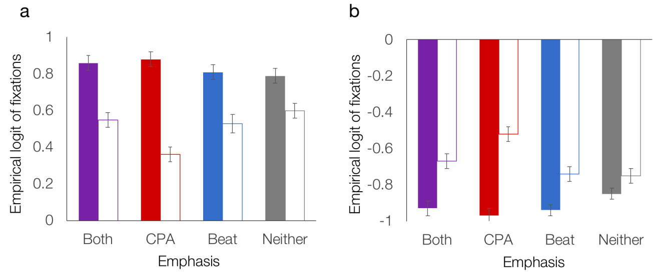

Figure 2: Fixations on (a) target and (b) competitor objects during the shape word period for color-contrast (filled) and both-difference (outlined) critical referring expressions by prominence cue (CPA, Beat, Both, Neither).

We also observed a continuation of the baseline color-contrast preference: There was a higher likelihood of fixating target objects and a lower likelihood of fixating competitor objects during color-contrast than both-difference critical referring expressions. Critically, this preference was qualified by interactions with beat gesture for target fixations (B = 0.13, SE = 0.06, t = 2.18, p = .03) and CPA for competitor fixations (B = 0.26, SE = 0.10, t = 2.51, p = .02; see Fig. 2). A simple-effect analysis revealed a greater likelihood of fixating target objects during color-contrast than during both-difference critical referring expressions when beat gesture was present (color: M = 0.84, SD = 0.80; both: M = 0.54, SD = 0.87; t = -2.01, p = .001) than when it was absent (color: M = 0.83, SD = 0.75; both: M = 0.48, SD = 0.86; t = -3.55, p = .047), indicating that beat gesture facilitated online resolution of contrastive critical referring expressions. Another simple-effect analysis revealed a greater likelihood of fixating competitor objects during both-difference than color-contrast critical referring expressions when CPA was present (both: M = -0.59, SD = 0.76; color: M = -0.95, SD = 0.76; t = 4.49, p < .001) than when it was absent (both: M = -0.75, SD = 0.84; color: M = -0.89, SD = 0.69; t < 1), indicating that CPA contributed to incorrect contrastive interpretation of non-contrastive critical referring expressions. Together, these results indicate that beat gesture and CPA serve as cues to contrast during online reference resolution in spoken discourse. Further, the absence of any significant Gesture x Accent interactions indicates that the effects of these cues are additive rather than interactive.

Discussion

Consistent with our predictions, the results indicate that the effects of beat gesture and CPA vary by contrast type during online reference resolution in spoken discourse. Specifically, beat gesture encouraged fixations on target objects during resolution of color-contrast critical referring expressions, confirming that beat gesture can convey contrast effectively. Moreover, CPA encouraged fixations on competitor objects during resolution of both-contrast critical referring expressions, indicating that it acted as a "garden path" resulting in an incorrect contrastive interpretation. Together, these results indicate that beat gesture and CPA each encourage contrastive resolution of referring expressions during online spoken discourse processing. By providing the first evidence that beat gesture facilitates online resolution of contrastive referring expressions, the results of the current study build upon previous findings that beat gesture (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018; Morett & Fraundorf, under review; Morett, Roche, Fraundorf, & McPartland, 2018) and CPA (Fraundorf et al., 2010; 2012; Lee & Fraundorf, 2016; Lee & Snedeker, 2016; Sanford, Sanford, Molle, & Emmott, 2006) enhance processing and memory of contrastive information in spoken discourse.

Considered in conjunction with the separate interactions with contrast discussed above, the lack of significant interactions between beat gesture and CPA indicates that these cues exert independent, additive effects on online contrastive reference resolution. This finding differs from work on discourse memory, which has shown interactive effects of beat gesture and CPA (Kushch & Prieto, 2016; Llanes-Coromina et al., 2018; Morett & Fraundorf, under review). Although the reasons for this difference are not entirely clear, one possibility is that separate effects of these cues on contrastive information processing in spoken discourse interact during storage or retrieval, leading to the interactive effects observed in studies of offline processing. Another possibility is that effects of these cues that appear separate in the short-term become interactive in the long-term. Future research should distinguish between these possibilities by introducing a delay during which recollection either is or is not required for discourses containing contrastive information in which these cues are varied.

It is worth noting that beat gesture and CPA affected target and competitor object fixations during the shape word but not the color word period, indicating that these cues affect resolution—but not anticipation—of referents. The timing of the effects of these cues is consistent with some previous work examining the effect of CPA on reference resolution (Ito & Speer, 2008), but is inconsistent with other work examining this same phenomenon (Kurumada et al., 2014; Watson, Tanenhaus, & Gunlogson, 2008) as well as work examining the effect of representational gesture on reference resolution (L. B. Silverman et al., 2010). While the reasons for the absence of effects of beat gesture and CPA on reference anticipation in the current study are not entirely clear, one possibility is that these cues may have elicited a processing cost, increasing reference resolution latency. This possibility is consistent with pupillometry data from the current study (Morett, Roche, Fraundorf, & McPartland, 2018), which indicates that the combination of beat gesture and CPA increases cognitive load during reference resolution. Alternatively, fixations to the video may have persisted during color word processing, reducing fixations to target and competitor objects. This possibility is consistent with the results of an analysis of target fixations during the color word period in which the video was included as an interest area, in which the significant main effect of contrast observed during this interest period in the model excluding the video interest area was absent. Thus, both cognitive load and persistence of fixations on the video may have contributed to the timing of the effects of beat gesture and CPA on online reference resolution in the current study.

In conclusion, the results of the current study indicate that beat gesture and CPA exert independent, additive effects that facilitate online contrastive reference resolution during spoken discourse processing. As such, they provide the first evidence that beat gesture is interpreted as a cue to contrast during online spoken discourse comprehension.

Acknowledgments

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

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*, 92-103.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457-474.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling. *The OSU Research Foundation*, *3*, 30.

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes*, *25*, 1024–1043.

Esteve-Gibert, N., & Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, *56*, 850–864.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*, 367–386.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2012). The effects of age on the strategic use of pitch accents in memory for discourse: a processing-resource account. *Psychology and Aging*, *27*, 88–98.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151-171.

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., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, *66*, 265–284.

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.

Kuznetsova, A. Brockhoff, P.B., & Christensen R.H.B. *lmerTest: Tests for random and fixed effects for linear mixed effect models*. R package version 2.0-0. www.r-project.org (December 2018, date last accessed).
Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*, 335-342.
Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. *Language, Cognition and Neuroscience,* 1–13.
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: Int'l Speech Communication Association.
Ladd, D. R. (1996). *Intonational phonology*. New York: Cambridge University Press.
Lee, E.-K., & Fraundorf, S. H. (2016). Effects of contrastive accents in memory for L2 discourse. *Bilingualism: Language and Cognition*, 1–17.
Lee, E.-K., & Snedeker, J. (2016). Effects of contrastive accents on children's discourse comprehension. *Psychonomic Bulletin & Review, 23,* 1589-1595.
Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26,* 1457–1471.
Llanes-Coromina, J., Vilà-Giménez, I., Kushch, O., Borràs-Comes, J., & Prieto, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology, 172,* 168–188.
McNeill, D. (1992). *Hand and mind.* Chicago: University of Chicago Press.
McNeill, D. (2005). *Gesture and thought.* Chicago: University of Chicago Press.
McNeill, D. (2006). Gesture: A psycholinguistic approach. In E. Brown & A. Anderson (eds.), *The Encyclopedia of Language and Linguistics* (pp. 58–66). Boston: Elsevier.
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. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98*, 834–853.
Morett, L. M. & Fraundorf, S. H. (under review). Beat gesture alters how pitch accenting affects discourse memory: Evidence from top-down use of talker expectations.
Morett, L. M., Roche, J.M., Fraundorf, S. H., & McPartland, J.C. (2018). Pupillometry and multimodal processing of beat gesture and pitch accent: The eye's hole is greater than the sum of its parts. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.
Myhill, J., & Xing, J. Z. (1996). Towards an operational definition of discourse contrast. *Studies in Language*, *20*, 303-360.
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. Presented at *5th International Conference on Speech Prosody*, Chicago, IL.
Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, *137*, 172-180.
Sanford, A. J., Sanford, A. J., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*, 109–130.
Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the Third International Conference on Spoken Language Processing* (pp. 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.
Vilà-Giménez, I., Igualada, A., & Prieto, P. (in press). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Developmental Psychology.*
Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: An ERP study. *Neuropsychologia*, *51*, 2847–2855.
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.
Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*, 367–392.