BUCLD 38 Proceedings
To be published in 2014 by Cascadilla Press
Rights forms signed by all authors

Directing Toddler Attention: Intonation Contours and Information Structure

Jill C. Thorson and James L. Morgan

1. Introduction

From birth, infants are sensitive to native language rhythm and pitch patterns (Nazzi, Bertoncini & Mehler, 1998; Nazzi, Floccia & Bertoncini, 1998). From the onset of speech production at around one year of age, babies can approximate adult-like intonation contours (Prieto & Vanrell, 2007; Chen & Fikkert, 2007; Frota & Vigário, 2008) and align these contours with felicitous semantic and pragmatic intentions (Prieto, Estrella, Thorson & Vanrell, 2012). However, little research has been conducted on the early *comprehension* of contours as they reflect information status. Previous research suggests that toddler attention to referents is mediated by both intonation and information structure in discourse (Grassmann & Tomasello, 2010). In turn, attention to referents is essential for making the correct word-to-object or word-to-action mappings necessary for early word learning (Grassmann & Tomasello, 2007). The motivation for our study is to investigate how American English-acquiring 18-month-olds are guided by mappings from intonation to information structure during on-line reference resolution in discourse.

Previous research by Grassmann and Tomasello (2010) claimed that German-acquiring two-year-olds attend to a referent of a familiar word if and only if the word is both *stressed* and *new* to the discourse context. Their experiment consisted of three conditions where the target referent could be introduced in a brief discourse context with either (a) *stress* only, (b) *newness* only, or (c) both *stress* and *newness*. Grassmann and Tomasello used a live speaker to present stimuli to each subject in order to ensure a level of intentionality on the part of the speaker and measured looking time and pointing to a referent as their dependent variables. Children in their experiment looked equally as long at the distractor referent as at the target when it was *new* only (not stressed) or *stressed* only (not new). Critically, they found reliable looking to the target referent in the third condition only, where it was both *new* and *stressed*.

* Authors' affiliation: Brown University, Department of Cognitive, Linguistic and Psychological Sciences. Correspondence email: jill_thorson@brown.edu. Special thanks to the subjects, their families, Laura Kertz, Pilar Prieto, Lori Rolfe, Megan Keough, Glenda Molina-Onario and the Infant Research Lab. This work was supported by NIH grant R01HD068501 to JLM.

Our experiment expands upon the work by Grassmann and Tomasello, introducing a number of methodological changes. First, we ask whether specific pitch movements more systematically predict patterns of attention to referents, rather than using one collapsed "stressed" category. Second, we add a deaccented condition to the design to act as a control against which the accented (or 'stressed') conditions may be compared. Third, because a live speaker may have produced varying intonation contours, we control for speaker variations and possible experimenter bias by using pre-recorded stimuli. This in turn allowed us to isolate the role of pitch in guiding attention while holding duration and intensity constant.

Instead of one 'stressed category', we are interested in how unique pitch accents affect toddler attention. We focused on two pitch accents in American English, the simple monotonal H* and the complex bitonal L+H*. Previous work in adult speech perception shows that the simple H* pitch accent can be associated with either new or contrastive information in discourse, while the complex L+H* accent is more typically associated with a contrastive interpretation (Watson, Gunlogson & Tanenhaus, 2008). Although we only manipulate pitch (F_0) in our experiment, we exploit these mappings in American English in order to test how these specific pitch accent movements interact with referential newness and givenness during early attentional processes. For this study, the three pitch type movements selected were a simple monotonal (~H*), a complex bitonal (~L+H*) and a deaccented control.

Information structure in this experiment is used to describe the simple dichotomy between *new* and *given* information. For each of the short discourse contexts presented, a referent is considered *new* if it has not been previously mentioned or seen by the participant. A referent is *given* if it has been previously mentioned and seen during the discourse context prior to the test phase.

Finally, past research has emphasized the role of social pragmatics and intentionality in achieving successful word learning (Tomasello & Akhtar, 1995; Tomasello, Akhtar, Carpenter & Tomasello, 1996). Importantly, our study tests outcomes when live interactions are removed from the experimental design and any degree of intentionality is only accessible through the utterances themselves and their corresponding prosody.

Our primary research question asks how the mapping between information structure and intonation guide toddler attention in a controlled discourse context. We expand upon past research in order to test how sensitive 18-month-old toddlers are to specific pitch accent movements in American English, depending on the discourse context. Our predictions are three-fold. First, we predict that newness will be sufficient to guide attention to a referent, regardless of pitch accent type. That is, even when a target is deaccented when introduced, the novelty of the referent will still attract the attention of the toddler. Studies of visual attention in the absence of discourse have consistently found that infants and toddlers prefer to look at novel patterns or objects, and we would expect this bias to influence looking during discourse as well (Fantz, 1964). Second, we predict that a semantically and pragmatically appropriate pitch accent will guide

attention to both *new* and *given* referents. As previously mentioned, the pitch accents that will be used in this study have been shown to guide attention to presentationally new (H*) and contrastive information (both H* and L+H*) (Watson et al., 2008). Consequently, these two pitch accents should guide attention to a contrastive referent, even if it is *given* in the discourse. Our third prediction is that the complex pitch accent will show more looking to the target referent than the simple pitch accent, specifically in the case where the referent is *given* in the discourse context. We expect such an increase in looking due to the more prominent status of the bitonal pitch movement as well as its similarity to pitch types often employed in child-directed speech. Our predictions depart significantly from what has been reported in past research for how intonation interacts with information structure during online reference resolution.

2. Method

2.1 Participants

Data were analyzed for 40 American English-acquiring 18-month-old toddlers (22 female). The age of participants ranged from 529 to 589 days, with a mean age of 550 days. Ten additional participants were discarded due to fussiness (8) or equipment malfunction (2). All participants were from Providence, RI, USA, and surrounding areas.

2.2 Stimuli

The same female speaker produced all target and distractor utterances. To create the three different pitch types, the speaker first produced carrier sentences with H* accents on the target words using careful (slow and clear) but not child-directed speech. The pitch contour of the target word only was then digitally manipulated in Praat in order to create the two accented versions, the simple monotonal and the complex bitonal (Boersma & Weenink, 2012). For the deaccented versions, deaccented productions were spliced into the same carrier phrase as the two accented types (*simple* and *complex*) and matched for duration and intensity. All test stimuli were resynthesized. Naïve listeners judged the resynthesized target speech sounds as sounding natural, although sometimes a bit faster than the rest of the utterance.

Six (C)CVC monosyllabic target words and 18 (C)CVC monosyllabic distractors were used in the procedure (see Table 1 for a full list of stimuli by trial). All target words were phonologically distinct within a trial and were primarily sonorant in nature. All of the target and distractor words used are commonly known by 18-month-olds. Previous knowledge of the words was confirmed by a vocabulary questionnaire completed by the caregiver. If the toddler was not familiar with a particular word, then that particular trial was eliminated during analysis (this was not a common occurrence and affected at most one trial per participant).

Table 1. *Stimuli for Experiment 1.*

Trial Number	Target Word	Distractors/Fillers
Practice	spoon	cake bear fish
1	ball	sock lamb cat
2	moon	dog shoe book
3	cow	pig tree duck
4	doll	bus cup sun
5	plane	star truck dress

3.3 Design and procedure

We used a 2x3-mixed design to test toddlers' responses to variations in information status and intonation, isolating the specific role that pitch plays in directing attention to *new* or *given* referents. The two independent variables were Information Structure and Pitch Type. Information Structure was manipulated within-subjects and consisted of two levels: *new* vs. *given.* Pitch Type was manipulated between-subjects and consisted of three levels: *deaccented, simple monotonal,* and *complex bitonal.*

Each trial consisted of a Context Phase and a Test Phase. The Context Phase consisted of two parts (or slides), and established the Prior Discourse Context needed for the introduction of the Test Phase (see Figure 1). During the Test Phase, the test utterance played and the *proportion of looking time (PLT) to the target* was collected using a SMI iViewX™ RED eye tracker. Importantly, the target item in the test slide was either *new* or *given* in the discourse context, as well as in contrast to a referent in the previous slide, making both the simple (~H*) and the complex (~L+H*) movements acceptable in this context.

There were five familiarization trials and two test blocks per condition. Each test block tested one of the Information Structure levels: one block for *new* and one block for *given.* Pitch Type was tested between-subjects. Each test block included one practice trial and five test trials. With two blocks per

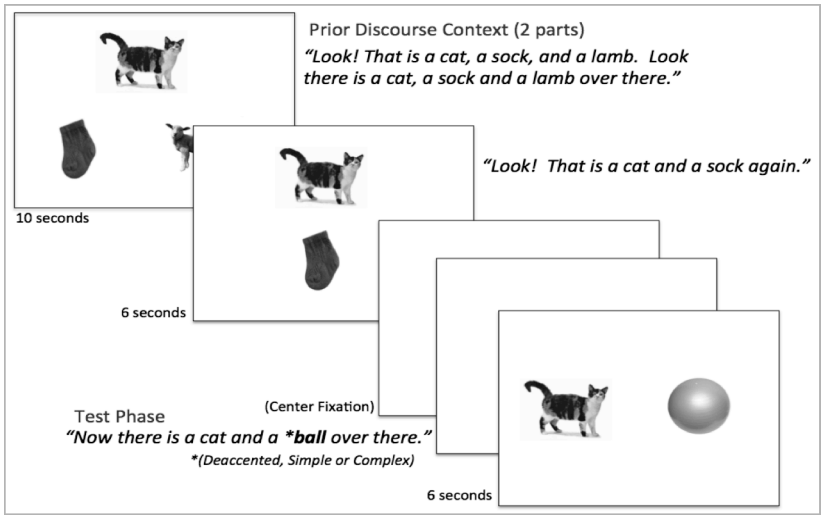

Figure 1: *Example new trial from Experiment 1. The target referent presented as deaccented, simple or complex depending on the condition. In a given condition, the target (e.g. 'ball') would have also been present in the first slide during Prior Discourse Context (e.g. in place of the 'lamb').*

condition, there were a total of two practice trials and ten test trials (twelve total trials) for each condition. Trial order within a block was randomized and block order was counter-balanced across participants. The location of the target items on the screen (left or right) was also counter-balanced within and across conditions.

3. Results

A 2x3 repeated measures mixed ANOVA showed a significant main effect of Information Structure ($F(1,37) = 26.10$, $p < .001$, $\eta_p^2 = .414$) and a significant main effect of Pitch Type ($F(2,37) = 4.27$, $p = .021$, $\eta_p^2 = .437$). The two-way interaction of Information Structure by Pitch Type approached but did not reach significance ($F(2,37) = 2.57$, $p = .09$).

Planned comparisons between groups revealed more looking to a target than a distractor when the referent is *new* to the discourse context, regardless of pitch accent type. In the deaccented condition, a paired t-test showed that there is significantly longer looking to the target over a distractor only when the referent was *new*, not when it was *given* ($t(13) = 4.86$, $p < .001$) (see Figure 2). Additionally, both types of accentuation (*simple* and *complex*) guide more looks to the target than to the distractor when the target referent was either *new* or *given* to the discourse context. In the *simple* condition, there was significantly longer looking to the target when the referent was *new* to the discourse context than when it was *given* ($t(12) = 2.42$, $p = .032$). This difference between *new* and

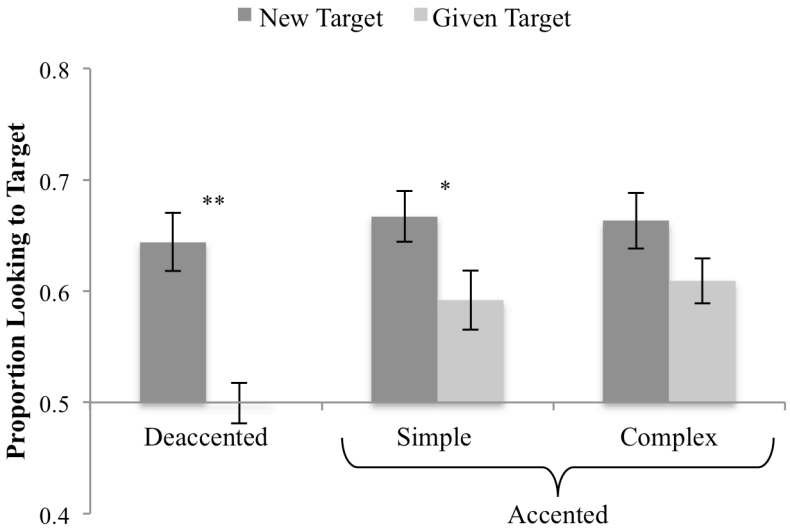

Figure 2: *Bar graph of proportion looking to a new or given target referent for each pitch type condition. Error bars show standard error. *: $p < .05$, **: $p < .01$.*

given target referents did not reach significance in the *complex* pitch type condition ($t(12) = 1.65$, $p = .125$). Crucially, there were more looks to the target referent than the distractor in all conditions except when the target is *deaccented* and *given* in the discourse.

4. Discussion

Contrary to Grassmann and Tomasello, but consistent with literature on visual attention, *newness* is sufficient to draw 18-month-olds' attention to referents in discourse, even without accentuation. A preference for the novel (or new) stimulus item over a familiar (or given) one suggests a more mature level processing by 18-month-olds (Fantz, 1964; Rose, Gottfried, Melloy-Carminar & Bridger, 1982). Toddlers prefer to look at the more prominent or salient item, where salience in this case is achieved through pitch movement and novelty effects.

As predicted, even in the case of a target referent that is *given* in the discourse context, both *simple* and *complex* pitch movements guide attention to the target. Thus, the presence of either *newness* or a *pitch accent* shifts attention to a target referent, regardless of pitch movement type. Again, this suggests that the more salient or prominent the stimulus item, the longer the toddler will look. For *given* information, attention is being driven by the pitch accent on the referent word. When there is both newness and a pitch movement, this results in

even greater looking to the target. This was especially evident in the *simple* pitch movement condition, where there was significantly longer looking to the target when it was *new* than when it was *given* in the discourse context.

We also predicted that there would be an increase in looking time to referents introduced with a *complex* pitch movement. While there was greater looking to *given* referents produced with a *complex* pitch movement over ones produced with *simple* movements, this difference was not significant. One explanation for why the *complex* pitch type results were not as robust as predicted is that the only acoustic cue available was F_0. If more acoustic correlates of intonation were present (i.e. duration and intensity), we anticipate that this would lead to an increase in saliency and thus greater looking in the case of the more *complex* pitch accent.

Our results reflect only pitch movement variations on the target word. Since duration and intensity were held constant across conditions, all effects were due to either the information status of the referent or its pitch type movement (F_0 movement only). An extension of this study will test what effects (if any) the other primary acoustic correlates of intonation have on toddler attention (e.g. duration, intensity). The question remains of how much each of the acoustic correlates of intonation weighs in directing attention.

In this experiment, the three pitch types analyzed were a simple monotonal pitch movement, a complex bitonal pitch movement, and a deaccented control. Critically, the methodological differences we introduced change the complexion of our results in comparison to previous research. Future work will extend analyses to other types of pitch accents and discourse contexts. Analyzing how higher-level components combine to direct attention to a referent in discourse informs early word recognition as well as early word learning. Ongoing work explores how information structure and contrastiveness interact with intonation to aid two-year-olds during a novel word learning task.

References

Akhtar, Nameera, Carpenter, Malinda, & Tomasello, Michael. (1996). The role of discourse novelty in early word learning. *Child development, 67*, 635-645.

Boersma, Paul, & Weenink, David. (2012). *Praat: A system for doing phonetics by computer*, available at http://praat.org.

Chen, Aoju, & Fikkert, Paula. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (ICPhS) (pp. 315-320). Saarbrücken, Germany.

Fantz, Robert L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668-670.

Frota, Sónia, & Vigário, Marina. (2008). The intonation of one-word and first two-word utterances in European Portuguese. Paper presented at the *XI International Conference for the Study of Child Language Conference* (IASCL), Edinburgh, Scotland.

Grassmann, Susanne, & Tomasello, Michael. (2007). Two-year-olds use primary sentence accent to learn new words. *Journal of Child Language, 14*, 23-45.

Grassmann, Susanne, & Tomasello, Michael. (2010). Prosodic stress on a word directs 24-month-olds' attention to a contextually new referent. *Journal of Pragmatics*, *42*(11), 3098-3105.

Nazzi, Thierry, Bertoncini, Josiane, & Mehler, Jacques. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nazzi, Thierry, Floccia, Caroline, & Bertoncini, Josiane. (1998). Discrimination of pitch contours by neonates. *Infant Behavior and Development, 21*(4), 779-784.

Prieto, Pilar, & Vanrell, Maria del Mar. (2007). Early intonational development in Catalan. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (ICPhS) (pp. 309-314). Saarbrücken, Germany.

Prieto, Pilar, Estrella, Ana, Thorson, Jill, and Vanrell, Maria del Mar. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, *39*(2).

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*, 704-713.

Tomasello, Michael, & Akhtar, Nameera. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development, 10*, 201-224.

Watson, Duane, Gunlogson, Christine, & Tanenhaus, Michael. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science, 32*, 1232-1244.