

# Considering the Source: Preschoolers (and Adults) Use Talker Acoustics Predictively and Flexibly in On-Line Sentence Processing

Sarah C. Creel (creel@cogsci.ucsd.edu)

Department of Cognitive Science, University of California-San Diego  
9500 Gilman Drive, La Jolla, CA 92093-0515 USA

## Abstract

The identity of the person talking is likely to constrain the things that they talk about. Adults can use talker acoustics to make on-line predictions about upcoming spoken material (Van Berkum et al., 2008). However, this cue to meaning may take time to learn. Do preschoolers consider who is talking when they are comprehending spoken sentences? I explored this question in two eye-tracked picture selection experiments. Experiment 1 showed that children and adults use vocal cues to talker identity in predicting the color of upcoming referents in spoken sentences. Experiment 2 showed that children and adults flexibly use acoustic cues to talker for first-person requests (“I want the square”) but reference to individuals for third-person requests (“Billy wants the square”). This suggests that children aged 3-5 years use who is talking to constrain the scope of reference in sentence processing, and know when this cue is likely to be useful.

**Keywords:** language development, talker identification, perspective-taking, spoken language processing

## Introduction

No two people sound alike. Some research indicates that this poses a challenge for language processing (Mullennix, Pisoni, & Martin, 1989; Nusbaum & Morin, 1992). However, it may also provide additional, helpful information to the comprehender. That is, knowing who is talking can provide useful information in processing spoken language. For instance, adult listeners make different predictions about upcoming information in a sentence depending on who is speaking it (Van Berkum, Van Den Brink, Tesink, Kos, & Hagoort, 2008), suggesting they have particular semantic associations with certain voice characteristics (e.g., a child’s voice vs. an adult’s voice). Thus, acoustic differences among talkers potentially have rich semantic associations (Geiselman & Crawley, 1983). But how long does it take the developing language learner to form and use these associations in comprehending language?

Children are sensitive to familiar perceptual information about talkers from a very early age. For instance, they are better at generalizing words between talkers with a familiar accent than between talkers with an unfamiliar accent (Schmale & Seidl, 2009). This suggests that they are sensitive to the acoustic details in the speech signal. Less is known about how much semantic information children glean from talker acoustics. We do know that children have less positive affective responses to (Kinzler, Dupoux, & Spelke, 2007) and associate unfamiliar clothing, and

dwelling with (Hirschfeld & Gelman, 1997) speakers who sound unfamiliar (they speak foreign languages). These studies suggest that children associate familiar-sounding speech with familiar objects and positive affect.

Beyond this, it is not clear whether children store more nuanced semantic information in relation to speech acoustics. This information might be somewhat difficult to learn for two reasons. First, children may be working to *ignore* talker-related acoustics to extract the attributes related to meaning (*dog* spoken by Mom still means the same thing as *dog* spoken by Dad, so why pay attention to irrelevant acoustic variation?). Second, knowing who is talking may only be useful what the person is referring to himself (“I really need a vacation”) and not when talking about things irrelevant to himself (“It’s raining outside”). That is, talker information may only be a reliable cue to meaning in a limited set of circumstances.

## Use of other non-phonemic acoustic attributes in comprehension

Though talker information has not been explored as an influence on children’s on-line sentence processing, recent studies on other non-phonemic acoustic cues—prosody and vocal affect—provide some hints about the potential of talker as a semantic information source during development. Children seem adept at processing prosodic information. Snedeker and Yuan (2008) showed that children were sensitive to a speaker’s intonational phrase boundaries in their interpretations of prepositional-phrase attachment. Ito, Jincho, Minai, Yamane, and Mazuka (2009) and Bibyk, Ito, Wagner, and Speer (2009) found that children as young as 6 years use pitch accent to constrain upcoming referents to a set of items contrasting on the pitch-accented dimension. These studies suggest that children attend to non-phonemic sound patterns that cue differences in meaning.

Children seem to have more difficulty processing cues to vocal affect. Morton and Trehub (2001) found that when vocal affect conflicts with verbal content (e.g. hearing “I get to eat ice cream” in a sad voice, or “My dog got hit by a car” in a happy voice), children cannot ignore the verbal content when reporting the talker’s affect (reporting the first sentence as sounding happy, and the second as sounding sad). Nonetheless, recent work by Berman, Graham, and Chambers (2009) using eye tracking, a more sensitive, implicit measure, suggests that children associate positive and negative vocal affect cues with positively- and negatively-valenced pictures (e.g. intact vs. broken dolls).

Children may be using these cues by making associations between sound properties and semantic attributes. For instance, pitch accent seems to semantically activate contrast sets. In the vocal emotion case, children might have associations between sad vocal cues and non-intact objects (Berman et al., 2009). This leaves open whether children are able to use non-phonemic acoustic information in the speech signal to make high-level inferences about the perspective of the talker.

In sum, children show some ability to glean semantic information from two non-phonemic acoustic information sources, prosody and vocal affect. Thus, one might expect that children would gain semantic information from non-phonemic acoustic cues to talker as well. However, it is not clear that children can go so far as to use it to invoke a particular talker's perspective.

### The current study

To explore children's ability to exploit talker information in comprehending spoken language, I presented child and adult participants with an eye-tracked picture selection task (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) directed by two fictional child talkers, Anna and Billy. Each child professed a preferred color (pink vs. blue), and then asked for pictures on screen (e.g. "the square"), which were always their preferred colors. The question of interest was whether children would visually fixate the pictures in the talker's preferred color over the non-preferred color pictures based on which talker they hear.

I deliberately chose gender-stereotyped color preferences, reasoning that capitalizing on children's preexisting knowledge would minimize working memory demands that might mask sensitivity. I also queried the children's own color preferences, to determine whether they were able to predict color preferences (i.e., make looks to the talker's preferred-color pictures) when those preferences did not match their own.

In Experiment 1, I considered whether children (as well as adults) were able to use talker information early in the sentence as a cue to upcoming referent color. That is, are they able to infer what shape the talker might request, given that the talker is Anna, who prefers pink? In Experiment 2, I assessed children's flexibility in using talker information by making talker identity on its own a useless cue to referent color. Specifically, each child talker asked for a shape for herself half of the time, and for the other child the other half of the time.

## Experiment 1

### Method

**Participants.** Children ( $n = 24$ , ages 3-5 years) were recruited from local day-care and preschool facilities, and participated in the study at their day-care/preschool location. They were given a small toy as a thank-you gift. An additional two children were excluded due to high error

rates (50% and 63%). Adults ( $n = 29$ ) were recruited from the University of California San Diego human participant pool, and received course credit for participation.

**Visual stimuli.** Pink and blue squares, triangles, circles, and five-pointed stars were constructed in Microsoft PowerPoint and saved as 200 x 200 pixel .jpg files. Scenes of Anna with pink objects (a tutu, a bed, bunny slippers) and Billy with blue objects (a truck, a baseball cap, a watergun) were 1024 x 768 pixel .jpg files.

**Auditory stimuli.** Two native southern-Californian university students recorded requests for shapes, and descriptions of Anna's and Billy's favorite colors, in child-directed English. Recordings were made in a sound-attenuated chamber and saved to .wav files on a computer. Each utterance was edited for clarity, saved to its own sound file, and normalized to 70 dB. Target word (e.g. "square") onset was at 1003 ms after the sentence began, on average.

**Procedure.** Each experiment had four brief phases. During each phase, sound was presented over high-quality headphones as visual stimuli were presented on an LCD monitor. First, each talker appeared, surrounded by three pink (or blue) objects, and stated his/her preferred color. The talker named each colored object in turn. Children were then tested in their ability to distinguish the colors: on eight trials, they saw two of the same shape and heard Anna (Billy) ask "Where's the pink (blue) one?" Children did not proceed until they answered at least 7 of 8 trials in a row correctly. This verified that they could distinguish the two colors, and further reinforced each talker's preference. The two favorite-color trials were then presented again. Finally, there was a 32-trial test phase where Anna and Billy each requested objects (stars, squares, triangles, or circles). On each trial, children heard (for instance) Anna saying

(1) Can you help me find the square?

On every trial, two pictures were pink, and two were blue. Each talker requested squares, triangles, circles, and stars equally often. In this phase, neither talker used a color term, referring merely to the shapes themselves. Each shape+color combination occurred equally often in each screen position across trials. Each talker spoke on 50% of trials.

Adults clicked the desired picture with a computer mouse. Children pointed to their desired responses, which were then mouse-clicked by an experimenter. The measure of interest was whether participants, before knowing what shape was to be requested, would visually fixate pink things upon hearing Anna's voice and blue things upon hearing Billy's voice.

**Equipment.** The experiment was run in Matlab using PsychToolbox3 (Brainard, 1997; Pelli, 1997) and interfaced with the eye tracker using the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002). Participants' eye movements were recorded by an Eyelink Remote eye

tracker (SR Research, Mississauga, ON) at 4-millisecond (ms) resolution. Offline, this was down-sampled to 50-ms resolution to enable easier processing.

## Results

Figure 1 suggests that both children and adults were visually fixating pictures of the talker’s preferred color well before the onset of the target word. To quantify this, I analyzed the data as follows. First, trials with erroneous responses (7% overall) were discarded. Then, a measure of color preference was calculated, which I will call the “color-look score.” This was the proportion of looks to the non-target picture of the talker’s preferred color, minus averaged looks to the two nonpreferred-color pictures. When this quantity was zero, listeners were not looking at pictures of either color more than the other. When it exceeded zero, listeners were looking more toward the talker’s preferred color. (Negative values would imply looks to the talker’s nonpreferred color, but this result did not occur in the current experiment.) Bear in mind that eye movements based on spoken material were most likely planned about 200 ms before they occurred, meaning that eye movements planned based on a signal at 1000 ms will show up around 1200 ms (Hallett, 1986).

An analysis of variance (ANOVA) was calculated on participants’ color-look scores in three 400-millisecond (ms) time windows, with Time Window (200-600, 600-1000, 1000-1400) and Age (child, adult) as factors. The only significant factor was Time Window ( $F(2,102) = 23.49, p < .0001$ ). Individual  $t$ -tests indicated that both children and adults had color-look scores greater than zero—that is, they were looking more to the talker’s preferred color—by 200-600 ms (children:  $t(23) = 2.27, p = 0.03$ ; adults:  $t(28) = 3.21, p = 0.003$ ), which was also significant at 600-1000 ms ( $t(23) = 4.49, p = 0.0002$ ;  $t(28) = 5.64, p < .0001$ ) and 1000-1400 ms ( $t(23) = 7.35, p < .0001$ ;  $t(28) = 5.99, p < .0001$ ). Thus, both groups seem to be adept at utilizing talker information to decide whose preferences to invoke.

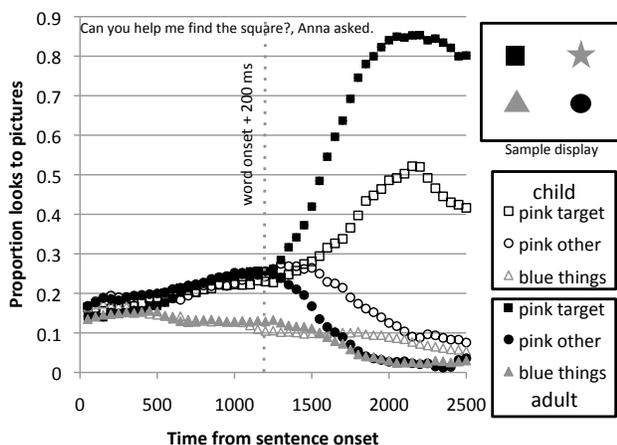


Figure 1: Adults’ (solid) and children’s (filled) looks to pictures in Experiment 1. Upper right inset: an example display where black=pink, gray=blue.

Note that children cannot be egocentrically fixating their own preferred color. If they were, then they should show no overall effect of the talker’s preferred color: pink looks on pink trials and pink looks on blue trials should cancel each other out. A more subtle version of this egocentricity hypothesis is that children only fixate the talker’s preferred color when it matches their own preferred color. This does not explain the results either: children whose preferred color matched neither talker ( $n = 12$ ) still showed above-chance looks to the talker’s preferred color at 600-1000 ms ( $t(11) = 3.75, p = 0.003$ ) and 1000-1400 ms ( $t(11) = 5.65, p = 0.0001$ ). This implies that children can use their knowledge of other individuals’ color preferences, even when different from their own, to constrain the domain of reference.

## Discussion

Both children and adults were able to use talker information early in the sentence to “predict” the color of the upcoming referent: they looked more at blue things when Billy began talking, and at pink things when Anna began talking. This verifies that, in a relatively simple situation, children use talker identity to constrain the referential domain of upcoming sentential material. Children showed looking effects equivalent to adults, suggesting that they are as able as adults to integrate talker information with verb information (Anna + want = pink, Billy + want = blue). This may depend on event knowledge that children have obtained through lifetime experience, or based on experimental conditions, but in either case, children are able to exercise this knowledge.

This experiment nicely demonstrates that children as well as adults are able to use talker characteristics to shape predictions of upcoming referents. One account of these data is that children and adults are using talker information to decide whose preferences to invoke to determine upcoming reference—they are constraining the domain of expected reference by talker. However, another explanation is that participants made a simple low-level audio-visual association between talker-related acoustic properties and color. That is, they associated the sound of a talker’s voice with pinkness or blueness, rather than using talker acoustics to access a representation of the talker as an individual with a color preference. On this latter account, they might look at pink things even if Anna were to say “Let me out of this cage” because her voice is associated with pink things.

Related to this issue is whether children are aware of contexts where talker information is even useful in real-world language processing. In particular, talker identity in the real world may only be useful for prediction when the talker is talking about himself. When the talker is talking about someone else—for instance, if Billy said that Anna wanted to see a particular shape—it would be disadvantageous to activate colors associated with Billy’s voice. This means that a smart listener would be able to use talker information in some (first person) situations, and ignore it in other (e.g. third person) situations. Presumably

adults do this readily, but it is unclear whether children do so.

## Experiment 2

Experiment 2 explored whether children and adults were able to use talker information to activate characteristics (i.e., color preferences) of each individual. The experiment was introduced as before, but now in the test phase each talker asked for a shape either for herself or for the other talker, followed by “Can you show me/him/her where it is?”:

- (2) Anna: I want to see the square. Can you ...
- (3) Billy: Anna wants to see the square. Can you ...
- (4) Billy: I want to see the square. Can you ...
- (5) Anna: Billy wants to see the square. Can you...

If children are learning low-level auditory-visual associations between talkers and colors, they should fixate pink things for (2) and (5) and blue things for (3) and (4). However, if they are learning information about *individuals*, then they may use talker information only in first-person cases, and use reference to Anna or Billy in third-person cases, to determine whose preferences to invoke. If so, they should look to the *agent’s* preferred color-pictures, looking at pink things in (2) and (3) and blue things in (4) and (5).

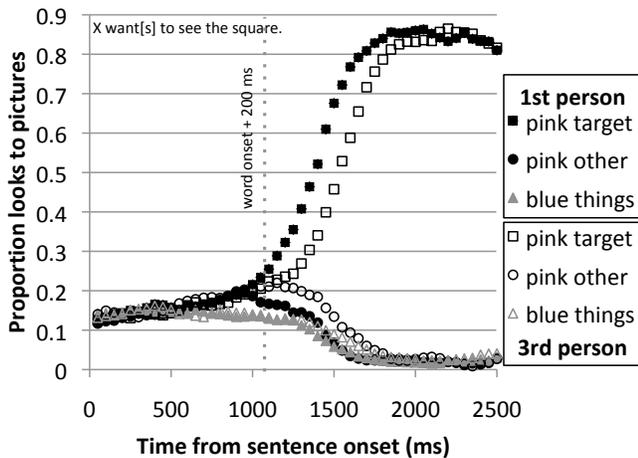


Figure 2: Adult fixations to targets and other pictures on 1<sup>st</sup>-person (circles) and 3<sup>rd</sup>-person (squares) trials.

## Methods

**Participants.** Children ( $n = 33$ ) and adults ( $n = 39$ ) were recruited as in Experiment 1. Two more children with extremely high error rates (34% and 44%) were excluded.

**Auditory stimuli.** A new set of spoken instructions were recorded by the same individuals as in Experiment 1.

**Procedure and Equipment.** These matched Experiment 1 in all respects.

## Results

Both adults (Figure 2) and children (Figure 3) seem to use talker information flexibly: when Anna is the agent of the sentence, they fixate pink things, regardless of whether Anna is the person talking. There were also somewhat later target fixations in the 3<sup>rd</sup>-person condition than in the 1<sup>st</sup>-person condition. While visually striking, this simply results from the 3<sup>rd</sup>-person sentences being slightly longer in duration than the 1<sup>st</sup>-person sentences (averaging 970 ms to word onset vs. 798 ms to word onset, respectively).

Error trials (5%) were discarded. Then, I conducted an ANOVA on color-look scores with Age (child, adult), Time Window (200-600, 600-1000, 1000-1400) and Person (1<sup>st</sup> person, 3<sup>rd</sup> person) as factors. This bore out the above observations. There was an interaction of Age x Time Window x Person ( $F(2,140) = 5.18, p = 0.007$ ), so individual ANOVAs were conducted for each Age. For adults, only Time Window was significant ( $F(2,76) = 10.3, p = 0.0001$ ), with color-look scores increasing over time. *T*-tests indicated that both 1<sup>st</sup>- and 3<sup>rd</sup>-person trials showed significant color looks at 600-1000 ms ( $t(38) = 2.13, p = 0.04$ ;  $t(38) = 2.73, p < 0.01$ ), and 1000-1400 ms ( $t(38) = 2.25, p = 0.03$ ;  $t(38) = 4.08, p = 0.0002$ ). For children, there was an effect of Time Window ( $F(2,64) = 23.48, p < .0001$ ), with color-look scores increasing over time, and a Time Window x Person interaction ( $F(2,64) = 3.36, p = 0.04$ ). *T*-tests comparing 1<sup>st</sup>-person and 3<sup>rd</sup>-person looks suggested nonsignificant differences in each time window (only 600-1000 ms approached significance,  $t(32) = 1.82, p = 0.08$ ). Regardless, both 1<sup>st</sup>- and 3<sup>rd</sup>-person color-look scores were significant at 600-1000 ms ( $t(32) = 2.22, p = 0.03$ ;  $t(32) = 4.84, p < .0001$ ) and 1000-1400 ms ( $t(32) = 8.11, p < .0001$ ;  $t(32) = 4.78, p < .0001$ ). This suggests that children, as well as adults, used the talker’s voice on 1<sup>st</sup>-person trials, but reference (the child’s name) on 3<sup>rd</sup>-person trials, to determine whose color preferences to use in constraining the referential domain. As before, results held for children ( $n = 18$ ) whose favorite colors were neither pink nor blue.

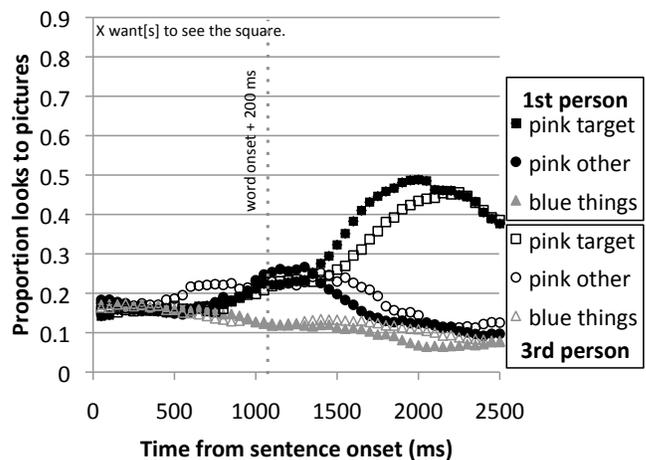


Figure 3: Child fixations to targets and other pictures on 1<sup>st</sup>-person (circles) and 3<sup>rd</sup>-person (squares) trials.

## Discussion

Children and adults in Experiment 2 succeeded at predicting the agent's color preference. That is, they made more visual fixations to shapes of the *agent's* preferred color on both first-person ("I want") and third-person ("Anna/Billy wants") trials. This implies that they use talker acoustics not just as a low-level auditory-visual association, but as a source of information about a participant in an action. Thus, children as well as adults can use non-phonemic acoustic information to activate information about an individual, and then infer the likely referential domain for that individual.

### General Discussion

Two experiments suggest that children are able to use their knowledge about particular talkers to constrain the domain of upcoming referents. In Experiment 1, listeners were instructed that Anna liked pink things, and Billy liked blue things. They then heard Anna and Billy request shapes of their preferred color. Both children and adults made more visual fixations to the shapes of the talker's preferred color than of the talker's nonpreferred color. This suggested that children were able to identify the talkers and use their individual preferences to constrain on-line interpretation of the request.

However, an equally good explanation was that children had associated female voice characteristics with pinkness, and male voice characteristics with blueness, a low-level auditory-visual cue correspondence rather than knowledge of an individual's preferences. Experiment 2 ruled out this explanation: listeners again heard Anna and Billy requesting shapes, but half the time, each talker requested a shape for the other talker. This meant that only on first-person trials ("I want") was talker a useful predictor, while on third-person ("Anna wants") trials, it was a misleading predictor. Impressively, children and adults were both able to use talker information on first-person trials, and proper nouns on third-person trials, to infer the identity of the sentential agent. That is, they always showed a visual fixation preference toward the *agent's* preferred-color shape, even when the agent was not the talker. This implies that, in a relatively simple task, children are able to use talker information selectively (only on first-person trials) to infer the identity—and thus the color preferences—of the agent.

### Implications for development of language processing

This research adds to the existing literature on cue integration in spoken language processing. Specifically, this work demonstrates that, in addition to prosody and vocal-emotional cues, non-phonemic acoustic cues related to talker can also be used to constrain processing on-line fairly early in life. This suggests excellent facility on the part of children to use non-phonemic acoustic cues to talker identity to understand the situation described by a sentence. This work is similar to adult research by Van Berkum et al. (2008), in which listeners showed a larger semantic

mismatch potential (N400) when the talker's identity and the action described were incongruous (e.g. a young child saying "I like to drink a glass of wine") than when they were congruous (an adult saying the same sentence). The current work suggests that preschool-aged children are similarly able to use talker acoustics to calculate likely (and unlikely) referents.

The current work, as well as Van Berkum's, fits nicely with a perspective on language processing (Kamide, Altmann, & Haywood, 2003; Bicknell, Elman, Hare, McRae, & Kutas, 2008) in which comprehenders use any available linguistic and nonlinguistic cues to construct event representations on-line. Acoustic information linked to talker identity is apparently useful in constructing event representations. Moreover, it is a robust enough cue that preschool-aged children can use it rapidly on-line (see Bates & MacWhinney, 1987; Snedeker & Trueswell, 2004 for further discussion of cue robustness and development).

Perhaps the most unique contribution of this study is the implication that children are using talker acoustics to infer *properties of individuals*, or at least of groups of individuals. That is, children are able to encode that Anna and Billy have particular color preferences, even when Anna and Billy have different preferences than the children themselves. As demonstrated in Experiment 2, this does not seem to be a simple auditory-visual association between Anna's voice (or female voices) and pink, and Billy's voice (or male voices) and blue, but an association with Anna and Billy as entities who have different preferences for color.

### Remaining questions

One obvious question is how much of children's ability to use talker information in this task is subserved by children's long-term knowledge of gendered color preferences. A quick visual search of major toy retailers' products confirms strong tendencies for female toys to be pink (or purple), and for male toys to be blue (or a number of other colors, but not pink). Thus, children's use of talker information here could be due to a lengthy learning process through exposure to gender-stereotyped objects in their environments. On the other hand, children might readily associate idiosyncratic preferences with particular individuals. If so, then children should also be able to use learned, non-gender-stereotyped color preferences to constrain on-line language processing.

An experiment in progress addresses this question, using black and white as the preferred colors. Only one child (1.5%) in Experiments 1 and 2 reported black as his favorite color, and none reported white, suggesting that children have little experience or gender-preference information for black and white. Further, color preference is counterbalanced across talker gender. With 15 child participants so far, there are robust looks to talkers' preferred colors. This suggests that neither conformance to a gender-stereotypical color mapping nor long-term learning is necessary for children to be able to use talker information predictively. However, talker gender itself may still be an

important social anchor point for encoding talker preference.

Another question is how subtle children are in their appreciation of talker information. Are they as keen in their perceptions as adults? If not, how do they differ from adults? Direct comparisons may be limited somewhat by children's level of social knowledge relative to adults—adults may only seem more adept at using talker cues because they have more subtle knowledge of social variation.

Finally, it is unknown how semantic knowledge based on talker characteristics relates to talker-specific perceptual facilitation of word-forms (e.g. Goldinger, 1996; see also Creel, Aslin, & Tanenhaus, 2008). Does talker-specific perceptual information covary with semantic usefulness? Despite these remaining questions, though, the current research forms a solid basis for further explorations of children's sensitivity to talker as a cue to meaning.

## References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Bates, E. A., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Berman, J., Graham, S. A., & Chambers, C. (2009). Preschoolers' appreciation of vocal affect as a cue to a speaker's intentions. Paper presented at *Boston University Conference on Language Development 34*, Boston, MA.
- Biby, S., Ito, K., Wagner, L., & Speer, S. (2009). Children can use contrastive pitch accent in on-line processing. Paper presented at *Boston University Conference on Language Development 34*, Boston, MA.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2008). Online expectations for verbal arguments conditional on event knowledge. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2220–2225). Austin, TX: Cognitive Science Society.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Cornelissen, F.W., Peters, E., & Palmer, J. (2002). The EyeLink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments & Computers*, 34, 613–617.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, 108, 633 – 664.
- Geiselman, R. E., & Crawley, J. M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning & Verbal Behavior*, 22, 15–23.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166 – 1183.
- Hirschfeld, L. A., & Gelman, S. A. (1997). What children think about the relationship between language variation and social difference. *Cognitive Development*, 12, 213 – 238.
- Ito, K., Jincho, N., Yamane, N., Minai, U., & Mazuka, R. (2009). Use of emphatic pitch prominence for contrast resolution: An eye-tracking study with 6-year old and adult Japanese listeners. Paper presented at *Boston University Conference on Language Development 34*, Boston, MA.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Science*, 104, 12577 – 12580.
- Morton, J. B., & Trehub, S. (2001). Children's understanding of emotion in speech. *Child Development*, 72, 834–843.
- Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365–378.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437 – 442.
- Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, 12, 583–601.
- Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49, 238–299.
- Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58, 574–608.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580 – 591.