

Children’s Learned Associations with Voice: Perspectives on Children’s Speech Perception in Language Acquisition

Nicholas P. Moores, Kevin B. McGowan, Meghan Sumner,
& Michael C. Frank

Departments of Linguistics and Psychology, Stanford University

1 Abstract

Speech perception provides an avenue with which to study children’s burgeoning ability to comprehend social information in the speech stream. The speech signal contains not only phonemic information necessary for word recognition but abundant indexical information about the talker, including talker gender (Perry, Ohde, & Ashmead, 2001; Goldinger, 1998; Johnson et al., 1999; Johnson, 2006). A growing body of work has begun to investigate the effects of listeners’ associations between talkers’ social identity and the phonetic cues in talkers’ voices, demonstrating that by adulthood listeners use talker information to make social inferences about a talker’s likely behavior (Van Berkum et al., 2008), especially when they expect talker identity to be useful or find it to be a reliable cue (Creel, Aslin, & Tanenhaus, 2008). Recent work suggests that children use acoustic cues to talker identity to constrain comprehension of spoken language (Creel, 2012), though the way in which children learn to integrate social knowledge with information from talker voice remains poorly understood. In this paper, we test the hypothesis that children are able to disambiguate between objects with gendered associations (a men’s pair of gloves and a women’s pair of gloves, say) based on talker voice. We explore children’s use of talker indexical information to infer speaker meaning through an experiment in which children interact with a web page on an iPad. Over 24 trials, children ages 3-5 were shown series of four images and asked by talkers to find one of the objects by clicking on it. During half of the trials, children heard a male talker’s voice, and during the other half they heard a female talker’s voice. In addition, half of the trials were non-competitor trials

in which there was only one image of the talker’s referent (hearing a man’s voice and seeing a man’s glove, say). The other half were competitor trials in which the target image competed with a variant which would stereotypically belong to a speaker of the opposite gender (hearing a man’s voice but choosing between both a man’s glove and a woman’s glove). Children’s reaction time between the utterance of the target word and their click on an image was logged, as was their choice of image. Our data suggest that by the age of 5, children regularly integrate phonetically-cued socially indexical talker information with their social knowledge of speaker characteristics to guide their interpretations of speaker meaning in a real world paradigm. From ages 3 to 5, the ability to disambiguate increases significantly. These results suggest not only that children make use of socially-nuanced talker-specific acoustic information by a young age, but reveal a robust understanding of gender stereotypes that guide their daily interactions with interlocutors and may bear on their linguistic and social development.

2 Introduction

2.1 Gender Information in Voice

The speech stream is rich with both phonetic information necessary for word recognition and talker information necessary for conveying and perceiving talker identity (Perry, Ohde, & Ashmead, 2001; Goldinger, 1998; Johnson et al., 1999). Contrary to early work in speech perception that posited that talker-specific acoustic information is normalized to give way to the abstract exemplars of phonemes necessary for listeners to correctly perceive words in their language, a substantial flurry of more recent work suggests instead that listeners use cues to talker identity in very early stages of speech signal processing (Remez et al., 1997; Johnson, 2005; Creel, Aslin, & Tanenhaus, 2008; Creel, 2012), if not immediately (Van Berkum et al. 2008). Work from the past twenty years has established speech perception as a rather talker-contingent process (Nygaard, Sommers, & Pisoni, 1994) where listeners rapidly integrate acoustic cues relaying phonemic information about speaker message and social information about talker voice together at the same time. It has been argued that these two flows of information are not only parallel but interactive, as social information in voice can heavily influence listeners’ perception of speakers’ words, phones, and discourse (Sumner, et al., 2013).

One of the most salient ways that interlocutors make use of talker information in voice is with respect to gender. Although the ways in which speakers perform gender is largely idiosyncratic to speech communities (Johnson, 2006), individuals

across speech communities and vernaculars have strong expectations and guidelines about how men and women ought to talk in accordance with their gender. Accordingly, there is a large social component to the realization of gender in voice, leading to voice differences between different genders of speakers that are larger than would be expected if this difference were due to biological factors alone (Johnson, 2005). That is, speakers perform gender in line with their community’s expectations for how males and females should talk (Johnson, 2005, 2006). Evidence from previous studies suggests that these expectations are quite robust, with adult listeners being able to correctly categorize voices as male or female even when acoustic correlates of gender such as fundamental frequency have been reduced or removed altogether (Fellowes, et al., 1997). Listeners are able to be convinced of the gender of the speaker of a voice even when the speech signal is ambiguous or not strongly stereotyped (Johnson, 2006). Expectations about speaker gender and speaker voice will cause listeners to adjust their online interpretation of speaker’s utterances (Johnson, 1989), even leading adults to perceive the same word differently if they are told it is spoken by a man as opposed to a woman. For example, in some communities the acoustic boundary between the vowels in *hood* and in *hud* are much different for male and female speakers (Johnson, 2005). Listeners will also adjust their expectations for fricative duration and frequency for speakers of different genders in accordance with gender stereotypes that exist in their communities (Strand & Johnson, 1996). Accordingly, listeners see a processing benefit when speaker voice is in line with their community’s guidelines and expectations for performance of gender, and see a corresponding detriment when speaker voices are ambiguous with respect to gender or are in violation of the stereotypes prescribing proper male and female acoustic phonetic behavior (Strand 1999a, 1999b). These data suggest a powerful role for gender in conveying identity. This can be seen in other modalities as well, as adults struggle to ignore identity information while attempting a visual sex categorization task (Bulthoff, 2009), and adults have been shown to categorize faces more accurately and more quickly by sex than by race (Contreras et al., 2013).

So what makes voice gendered? That males and females differ in their acoustic productions is well-known, with early experiments demonstrating significant differences in fundamental frequency and in formant structure between male and female speakers (Peterson & Barney, 1952). Much of this difference though is likely due to the fact that gender is performed. Differentiation of voice characteristics along a gendered dimension appears quite early in development, with children’s voices first differing by gender with respect to vowel formant frequencies (but not f_0) at the age of four, and then differing with respect to both vowel formant frequencies and fundamental frequency by the age of twelve. Even so, adults are able to reliably

distinguish whether four-year-old children are male or female based only on their voice (Perry, et al., 2001).

2.2 Children’s Use of Gender Information and Talker Information in Voice

Additional work has established that gender is a salient perceptive cue even for young children. Children are able to discriminate individuals by gender from a very early age, as children as young as 6 months are able to classify male and female voices (Miller, 1983), and by 8 months, children are able to match faces and voices together by gender (Patterson & Werker, 2002). By the age of 7, children are able to perform as well as adults in facial recognition tasks classifying individuals by internal facial structure (Wild, et al., 2000). As children encounter more tokens of male and female speech in their linguistic communities they are able to develop richer expectations about what speakers of different genders ought to sound like, and about what they are likely to say (Andersen, 1990). Children’s expectations about gender as a whole seem to be largely shaped by socialization through widespread cultural stereotypes (Greenwald & Banaji, 1995; Philips, et al., 1987) and by their own linguistic interactions with their parents (Bellinger & Gleason, 1982; Gleason, 1975; Greif, 1980), with children’s understanding of these stereotypes being strong by the preschool years (Andersen, 1990). Elaine Andersen’s work examining children’s use of social registers in role play demonstrated that when children between the ages of 4 and 7 were asked to enact parents’ speech, their pitch was significantly deeper when they were playing as fathers than as mothers. In addition, when they spoke as a father they modified their language by lowering their pitch, speaking more loudly (sometimes yelling), and using back and lower vowels. When they pretended to be mothers they spoke in higher pitch, used exaggerated intonation contours, and chose stereotypically female vocabulary. In their discourse, pretend fathers spoke about work, business meetings, how tired they were from working at the computer, and how they had to fire their secretary. In the role of mother, they complained about being exhausted from their errands. These patterns seemed to hold whether or not the children’s own parents spoke about these topics or whether or not their parents’ occupations were in line with gender stereotypes. These data suggest a large influence of social stereotypes on children’s expectations with regards to gender by the preschool years. Beyond gender alone, children’s ability to use talker information in voice by this age appears to be quite strong (Creel, 2012).

Prior work on children’s comprehension of gender has focused either on children’s discrimination of gender in voice or on children’s understanding and use of sociolin-

guistic stereotypes about gender, but there has been little work linking the two. We can think of children’s expectations about speaker gender in voice in relation to their existing use of talker identity information in voice. Acoustic information linked to talker identity is useful for preschool-aged children when conducting social representations, and is a robust enough cue that they can use it rapidly on-line (Snedeker & Trueswell, 2004). Social information about talker identity present in the speech stream is accessed quickly during processing (Van Berkum et al., 2008), and is very strongly linked to processing of speaker message (Sumner, et al., 2013). This on-line integration of phonemic and identity information in the speech stream suggests that listeners’ expectations regarding realization of spoken identity information may play a powerful role in determining how that information is processed by listeners and how listeners believe speakers are likely to behave. Work by Sarah Creel has demonstrated that children augment their comprehension by using voice characteristics to develop and later access knowledge about individuals (Creel & Tumlin, 2011), and that children can use talker-specific acoustic information to learn preferences related to an individual even when that individual’s preferences are not buttressed by gender stereotypes (Creel, 2012). In addition, she has demonstrated that adult listeners can use the joint presence of phonemic and identity information in voice to constrain the domain of what a speaker is likely to say or want (Creel, 2012), especially when listeners expect or find talker identity to be a useful cue (Creel, Aslin, & Tanenhaus, 2008). Adult listeners have also been shown to use talker information to make inferences about a talker’s likely behavior (Van Berkum et al. 2008), however, little work has yet been done to investigate the extent to which children make these same inferences, or to investigate the ways in which children use their accumulated social knowledge during speech perception to guide their interactions with interlocutors. This past work also raises several questions about what we might expect when examining children’s use of talker information about whole classes of individuals, one such class being speaker gender. Although gender may be a salient cue to identity in voice, how soon do children start using information about speaker gender to infer speaker intention or speaker meaning?

2.3 The Present Study

Our work attempts to build on prior work examining children’s use of talker identity information in voice by examining the links that children already have between voice cues and social cues by the preschool years, and investigating how children use those linked cues to guide their interpretations.

I propose to examine the effect of gender-specific acoustic information on voice

on the inferences children make in conversation. Previous research has demonstrated that children remain sensitive to much phonetic detail and acoustic information previously thought to be generalized, and that they use this fine-grained information to construct social categories and to access lexical candidates that would seem appropriate given the speaker’s social identity. Preliminary data with adults has also demonstrated that in online processing of the speech stream adults use the gender- and age-specific acoustic characteristics retrieved from the speaker’s voice to attend to objects that are semantically associated with the speaker given the speaker’s age and gender (McGowan, Moores, & Sumner, in prep). This effect has not yet been researched in children, however. The goal of our study therefore is to replicate these findings in a sample in which the effects of talker-specific voice characteristics on attention to socially-graded objects has not yet been examined, young children. In this study I will examine whether children, who are themselves beginning to construct the social categories they will carry with them into adulthood, are yet able to use gender-specific identity information in voice to infer speaker preferences and intentions.

It is our hypothesis that children will be able to successfully link talker information in voice to the expectations about gender they have received through their socialization, and that they will use this linkage to construct inferences about speaker intention and preference that are in concordance with their socialization. If our hypothesis is supported, we would expect to see children improve in their performance in the critical task, so that scores for older children in the competitor condition should exceed those of younger children and should be significantly above chance.

3 Experiment 1

3.1 Method

Participants. Participants were children age 3-5 ($N = 72$) at Bing Nursery School and the Children’s Discovery Museum in San Jose, CA.

Stimuli. Our visual stimuli are 36 object pairs, rated by adult judges on Amazon Mechanical Turk as differing significantly in whether they are likely to be owned by a man or a woman. In each pair, one item was rated as being very likely to be owned by an adult male, not very likely to be owned by an adult female, and not very likely to be owned by a child. The other item was rated as being very likely to be owned by an adult female, not very likely to be owned by an adult male, and not very likely to be owned by a child. Difference scores were calculated for each pair, and the 36

pairs with the highest difference scores were chosen as the visual stimuli for this study.

Equipment. The experiment is run on an internet-connected iPad in Guided Access mode as a webpage. The experiment webpage on the iPad receives the conditions for that subject from the lab server and submits the results to a spreadsheet on the lab server.

Procedure. When the experiment begins, five colored dots appear on the iPad screen; the dots turn to x's when the subject clicks on them. The child subject must successfully click on all of them in order to proceed to the actual experiment; this ensures that the child understands how to properly click an iPad screen and reduces the chance of erroneous clicks to the screen.

Once the subject has clicked on all five dots, twelve trials ensue in random order, with four images in a 2x2 grid appearing in each trial. At the beginning of each trial, a speaker's voice plays, asking the subject to *find my X*, where *X* is the name of an object on the screen. The experiment is run in two blocks, with block order manipulated between subjects, such that half the subjects experience a block with a male speaker first, and the other subjects experience a block with a female speaker first. Noncritical trials consisted of there being one item on screen matching the speaker's description, accompanied by three distractor images. Critical trials consisted of two competing images appearing that match the speaker's description, accompanied by two distractor images. Crucially, the competing images in the critical trials differed along a gendered dimension, such that one item was rated as being very likely to be owned by an adult male (and not by an adult female), and its competitor was rated as being very likely to be owned by an adult female (and not by an adult male) by participants in an online survey. Each subject's trials are randomly selected from one of 12 lists of trials. The distribution of lists is balanced between subjects such that every item from the 36 object pairs appears as often as every other item, and every item appears as the target item as often as it is a competitor.

3.2 Results

If children are truly mapping social stereotype information from their developing social categories to inferred speaker preferences based on talker information in voice, then we would expect them to choose the target object on competitor trials significantly more often than chance. In line with this hypothesis, we saw that in competitor trials, four- and five-year old children chose the target object (the one that would belong to the speaker if speaker preferences were congruent with social

stereotypes) the majority of the time (4 year olds: $M = 61\%$, $SD = 49\%$ and 5 year olds: $M = 67\%$, $SD = 47\%$), whereas three-year old children did not choose the target object more often than chance ($M = 45\%$, $SD = 50\%$).

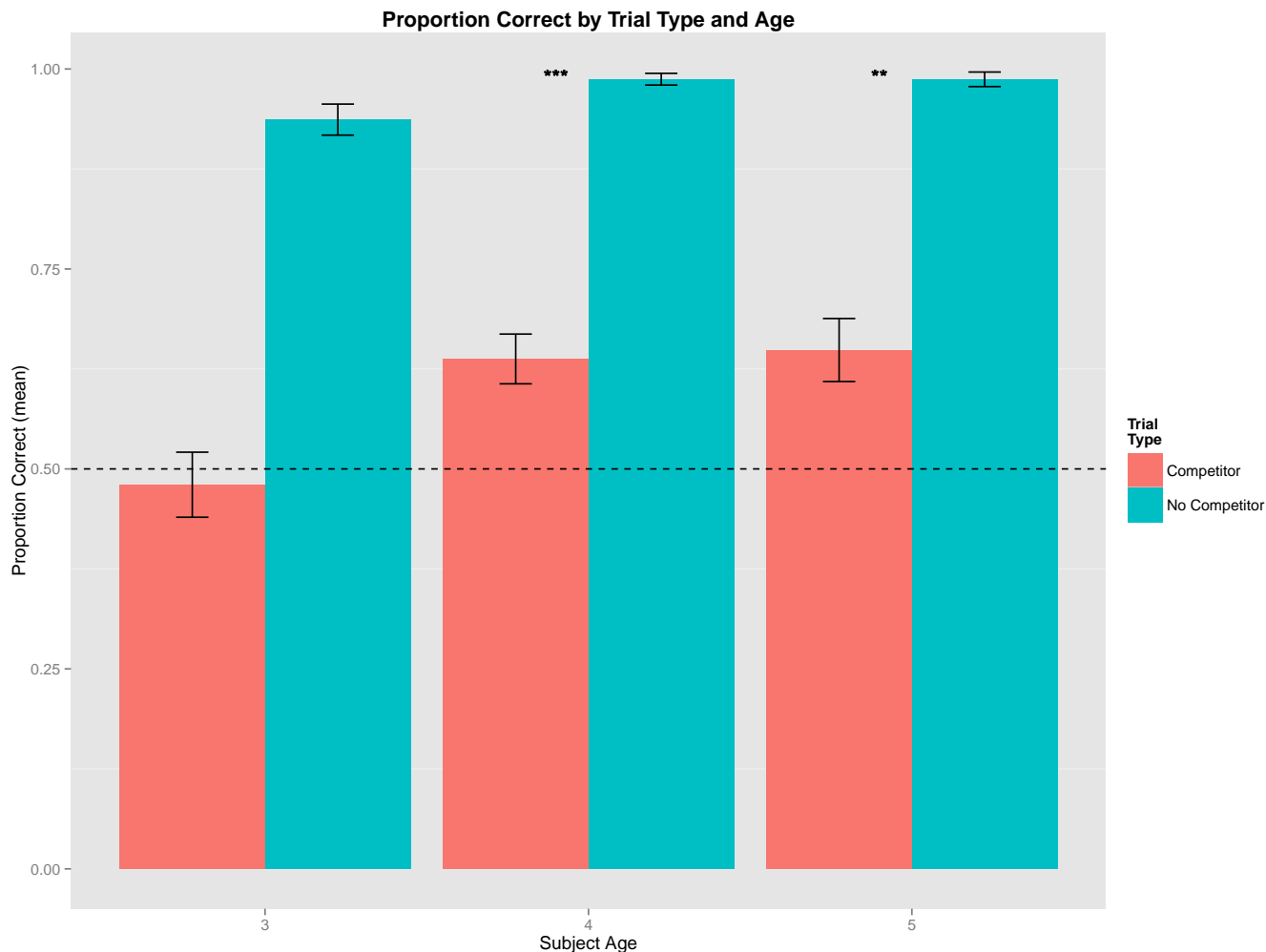


Figure 1: Mean Proportion Correct by Age

To quantify the reliability of these results, we fit a logistic mixed effects model (Gelman & Hill, 2006; Jaeger, 2008; Quen & Van den Bergh, 2008) to children's responses, with age group and condition as fixed effects, and with random effects of condition fit for each participant and each target item (Barr, Levy, Scheepers, & Tily, 2013; Baayen, Davidson, & Bates, 2008). The resulting coefficient estimates

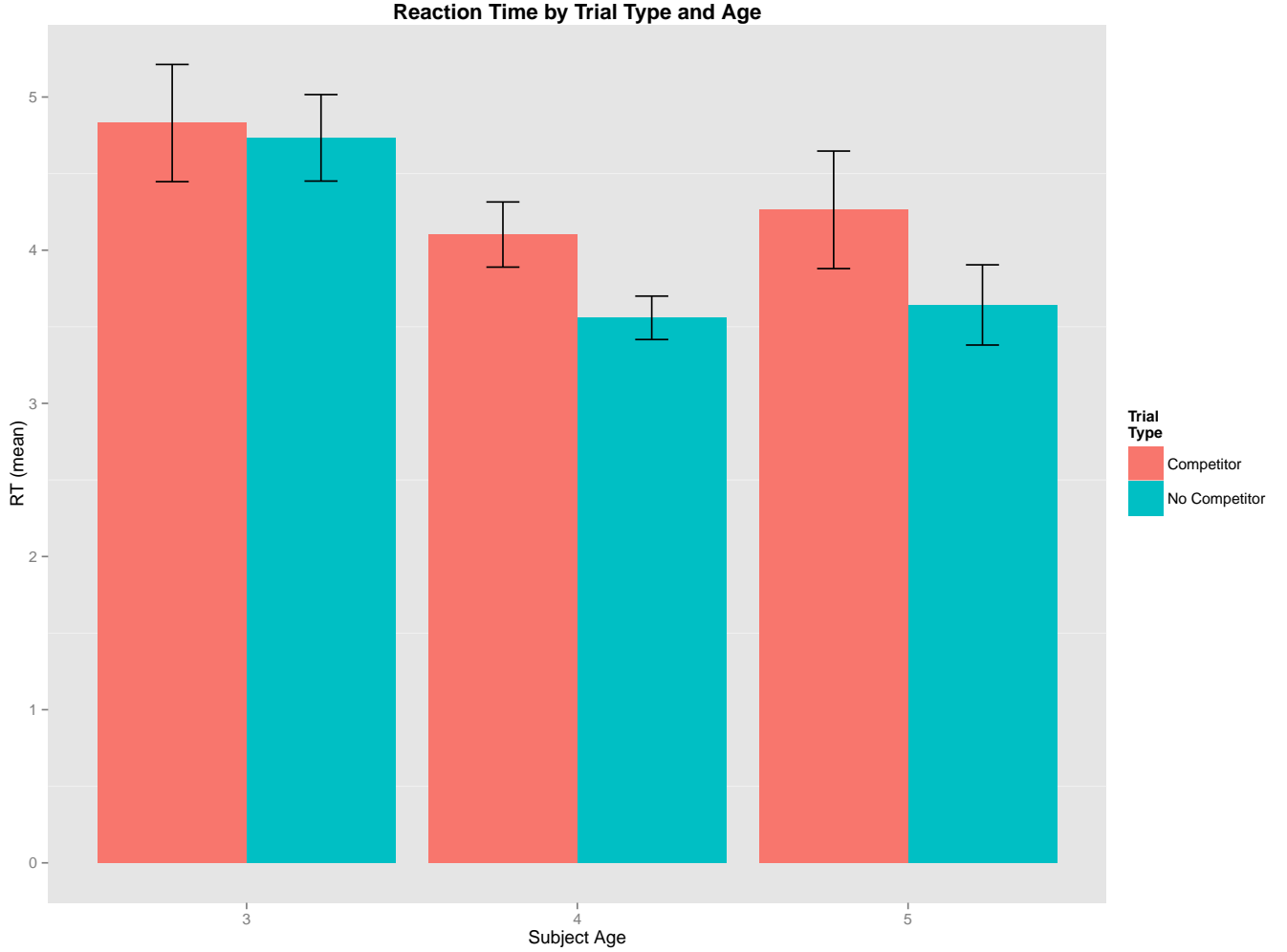


Figure 2: Mean Reaction Time by Age

suggested that three-year-olds (the reference level) were not above chance in their responding on critical trials ($\beta = -0.24$, $SE = 0.29$, $z = -0.819$, $p = 0.413$). There was a significant coefficient indicating higher performance on filler trials ($\beta = 3.22$, $SE = 0.31$, $z = 10.336$, $p < 0.001$). There were significant coefficients for the effect of age, suggesting that both four-year-olds ($\beta = 0.79$, $SE = 0.18$, $z = 2.794$, $p < 0.01$) and five-year-olds ($\beta = 1.10$, $SE = 0.32$, $z = 3.406$, $p < 0.001$) were above chance in their responses on critical trials, though five-year-olds were not significantly more accurate than four-year-olds ($\beta = 0.31$, $SE = 0.30$, $z = 1.031$, $p = 0.3$). There was

a significant coefficient for the effect of gender ($\beta = 0.76$, $SE = 0.24$, $z = 3.101$, $p = 0.002$), suggesting that female children performed significantly more accurately than male children.

3.3 Discussion

The data from this study suggest that children successfully link talker information in voice to developing forms of social categories, namely ideas and stereotypes about speaker gender. These data also suggest a developmental trajectory for children's linking of this information, as children are significantly above chance in their inference of a speaker preferring the object congruent with social gender stereotypes by the age of five, but at the age of three remain at or below chance.

These data leave several questions, however. Why do three-year-olds struggle with the task when four-year-olds do not? How might we interpret the significant effect of gender as a predictor of success in the task?

4 Experiment 2

Coming soon!

5 General Discussion

6 References

1. Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*, 113, 544-552.
2. Andersen, E. S. (1990). *Speaking with style: The sociolinguistic skills of children*. London: Routledge.
3. Andersen, E. S., & Johnson, C. E. (1973). Modifications in the speech of an eight-year-old as a reflection of age of listener. *Stanford Occasional Papers in Linguistics*, 3, 149-160.
4. Andersen, E. (1979). Register variation in young children's role-play speech. In *Proceedings of the Society for Research in Child Development*.

5. Aubrey, J. S., & Harrison, K. (2004). The gender-role content of children's favorite television programs and its links to their gender-related perceptions. *Media Psychology*, 6(2), 111-146.
6. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.
7. Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, 71(2), 230.
8. Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
9. Bellinger, D. C., & Gleason, J. B. (1982). Sex differences in parental directives to young children. *Sex Roles*, 8(11), 1123-1139.
10. Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2008). Online expectations for verbal arguments conditional on event knowledge. *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2220-2225). Austin, TX: Cognitive Science Society.
11. Brouwer, D., & de Haan, D. (1987). *Women's language, socialization and self-image (Vol. 4)*. Foris Publications.
12. Blthoff I, (2009). Sex categorization is influenced by facial information about identity. *Perception*, 38, ECVF Abstract Supplement, 78.
13. Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
14. Comstock, G., & Scharrer, E. The Use of Television and Other Screen Media. in Singer, D. G., & Singer, J. L. (Eds.). (2011). *Handbook of Children and the Media*. Sage publications.
15. Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PloS one*, 8(7), e69684.
16. 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.

17. Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, 65, 264-285.
18. Creel, S. C. (2012). Preschoolers' Use of Talker Information in On-Line Comprehension. *Child development*, 83, 2042-2056.
19. Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 498.
20. Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics*, 59, 839-849.
21. Gelman, A., & Hill, J. (2006). *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press.
22. Gleason, J. B. (1973). Code Switching in Children's Language. In Moore, T. (Ed.). *Cognitive Development and the Acquisition of Language* (Academic Press). pp. 169-167.
23. Gleason, J. B. (1975). Fathers and Other Strangers: Men's speech to Young Children. *26th Annual Georgetown University Roundtable* (Georgetown University Press). pp. 289-97.
24. 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.
25. Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105, 251.
26. Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1), 4.
27. Greif, E. B. (1980). Sex differences in parent-child conversations. *Women's Studies International Quarterly*, 3(2), 253-258.
28. Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28, 105-115.

29. Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.
30. Johnson, K. (1989) Higher formant normalization results from auditory integration of F2 and F3. *Perception and Psychophysics*. 46, 174-180.
31. Johnson, K. A., Strand, E. A., & D'imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 24, 359-384.
32. Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 363-389). Oxford: Blackwell.
33. Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics*, 34, 485-499.
34. Kempson, R. (2001). Pragmatics: Language and communication. In M. Aronoff & J. Rees-Miller (Eds.), *Handbook of linguistics*. Malden, MA: Blackwell.
35. Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29, 98-104.
36. Liberman, A. (1970). The grammars of speech and language. *Cognitive Psychology*, 1, 301-323.
37. Maccoby, E., & Jacklin, C. (1974). *The Psychology of Sex Differences*. Stanford: Stanford University Press.
38. McGillicuddy-De Lisi, A. V. & De Lisi, R. (Eds.). (2002). *Biology, Society, and Behavior: The Development of Sex Differences in Behavior*. Westport: Greenwood Publishing Group. Has even more gender stereotypes work and a variety of other perspectives on gender behavior.
39. Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, 6, 313-330.
40. Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42-46.

41. Nygaard, L. C. (2005). Perceptual integration of linguistic and non-linguistic properties of speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 390-413). Oxford: Blackwell.
42. Rideout, V. J., Vandewater, E. A., & Wartella, E. A. (2003). *Zero to six: electronic media in the lives of infants, toddlers and preschoolers*. Henry J. Kaiser Family Foundation.
43. Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents*. Henry J. Kaiser Family Foundation.
44. Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309.
45. Patterson, M. L. & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, 81, 93-115.
46. Perry, J. (1997). Indexicals and demonstratives. In C. Wright & R. Hale (Eds.), *Companion to the philosophy of language* (pp. 586-612). Oxford: Blackwell.
47. Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, 109, 2988-2998.
48. Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24, 175-184.
49. Philips, S. U., Steele, S., & Tanz, C. (Eds.). (1987). *Language, gender, and sex in comparative perspective* (No. 4). Cambridge University Press.
50. Quen, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425.
51. Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 651.

52. Rideout, V. J., Vandewater, E. A., & Wartella, E. A. (2003). *Zero to six: electronic media in the lives of infants, toddlers and preschoolers*. Henry J. Kaiser Family Foundation.
53. Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents*. Henry J. Kaiser Family Foundation.
54. Singh, S., & Murry, T. (1978). Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America*, 64, 81-87. Retrieved July 23, 2014, from <http://www.ncbi.nlm.nih.gov/pubmed/712004>
55. Sachs, J. (1987). Preschool boys? and girls? language use in pretend play. *Language, gender, and sex in comparative perspective*, 178-188.
56. Snedeker, J., & Trueswell, J. C. (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.
57. Sumner, M., Kim, S. K., King, E. & McGowan K. B. (2013). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in psychology*, 4.
58. Strand, E. A. & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In Dafydd Gibbon (Ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielefeld, October 1996*, (pp. 14-26). Berlin, Germany: Mouton de Gruyter.
59. Strand, E. A. (2000). Gender stereotype effects in speech processing (Doctoral dissertation, *The Ohio State University*).
60. Van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580-591.
61. Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters, part II: Recognition of rate-altered voices. *Journal of Phonetics*, 13, 39-52.
62. Wild, H., Barrett, S., Spence, M. J., O'Toole, A. J., Cheng, Y. D., & Brooke, J. (2000). Recognition and Sex Categorization of Adults' and Children's Faces:

Examining Performance in the Absence of Sex-Stereotyped Cues. *Journal of Experimental Child Psychology*, 77, 269-291.

63. Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11, 37-63.