6-2009

Statistical Frequency in Perception Affects Children's Lexical Production

Peter T. Richtsmeier
University of Kansas

LouAnn Gerken
University of Arizona, Tucson, Arizona, USA

Lisa Goffman
Purdue University, West Lafayette, Indiana, USA

Tiffany Hogan
University of Nebraska - Lincoln, thogan2@unl.edu

Statistical frequency in perception affects children's lexical production

Peter T. Richtsmeier,[a] LouAnn Gerken,[b] Lisa Goffman,[c] and Tiffany Hogan [d]

[a] Department of Psychology, University of Kansas
[b] University of Arizona, Tucson, Arizona, USA
[c] Purdue University, West Lafayette, Indiana, USA
[d] University of Nebraska–Lincoln, Lincoln, Nebraska, USA

Corresponding author — P. T. Richtsmeier, 1415 Jayhawk Blvd. Rm. 426,
Lawrence, Kansas 66045-7556, USA, tel 520 237-4485, email prichtsmeier@yahoo.com

Abstract
Children's early word production is influenced by the statistical frequency of speech sounds and combinations. Three experiments asked whether this production effect can be explained by a perceptual learning mechanism that is sensitive to word-token frequency and/or variability. Four-year-olds were exposed to nonwords that were either frequent (presented 10 times) or infrequent (presented once). When the frequent nonwords were spoken by the same talker, children showed no significant effect of perceptual frequency on production. When the frequent nonwords were spoken by different talkers, children produced them with fewer errors and shorter latencies. The results implicate token variability in perceptual learning.

Keywords: perceptual learning, speech development, talker variability, language acquisition, phonotactic probabilities

1. Introduction

It is now well known that both word learning and production speed and accuracy in childhood are affected by statistical frequency in the target language (e.g., Storkel, 2001, 2003). For example, Zamuner and colleagues demonstrated that the accuracy of two-year-olds' coda consonant production in CVC nonwords was influenced by the frequency of the preceding CV and VC bigrams in English words (Zamuner et al., 2004, 2005). Munson (2001) demonstrated a similar effect with three- to four-year-olds, who were asked to imitate CVCCVC nonwords. The duration and accuracy of children's productions for the words' medial consonant clusters was influenced by those clusters' frequency in English (cf. Edwards, Beckman, & Munson, 2004).

At least two mechanisms could be responsible for the effects of statistical frequency on young children's speech production. One is *articulatory practice*. That is, the likelihood that children will attempt to say a word with a particular sequence is higher for higher probability sequences. A

second possibility is *perceptual learning*. Numerous studies suggest that ambient language exposure allows infants to learn linguistic patterns, including phones (Maye, Werker, & Gerken, 2002), phonotactic probabilities (Chambers et al., 2003; Jusczyk et al., 1993, 1994; Saffran and Thiessen, 2003), and words (Saffran, Aslin, & Newport, 1996). These studies suggest that children's speech production might also depend on perceptual sensitivity to statistical information.

Two components of the perceptual learning hypothesis should be noted. First, it implies that learned patterns mediate between what is perceived and what is produced, suggesting some degree of abstraction (cf. Guenther, 2006; MacKay, 1989). Such mediation is thought to explain influences of perception on production in adult second language learning (Bradlow et al., 1997, 1999; Wang et al., 2003), but has not been shown for first language acquisition.

Second, perceptual learning may be influenced by factors other than the raw statistics of exposure (Johnson & Jusczyk, 2001). One such factor worth considering is phonetic variability. Work with adults suggests that listeners store fine-grained information about the phonetic signature

of a talker (e.g., Goldinger, 1996, 1998). In studies of infant language development, phonetic variability seems to facilitate word learning, with both talker voice (Houston, 2000) and affective variability (Singh, 2008)[1] leading to more robust word recognition. Our goal in the current work was to contrast the relative contributions of token frequency and phonetic variability to perceptual learning, and to contrast the articulatory practice and perceptual learning hypotheses as explanations for the effects of statistical frequency on child production. With respect to frequency, we manipulated the number of times children heard a nonword before being asked to produce it. We reasoned that if children differentially produced more vs. less frequent words, we would have evidence that perceptual frequency can drive production. With respect to variability, we exposed children to either identical acoustic tokens of a word or tokens spoken by different talkers. Finally, we varied the frequency in English of the medial consonant clusters of our words to examine how the frequency of newly learned items compared in different presentation conditions with frequency from previous experience (Edwards et al., 2004; Munson, 2001).

2. Experiment 1

The experiments manipulated the token frequency of CVC-CVC nonwords by presenting half of the words 10 times and half once. We refer to this as Experiment Frequency. In Experiment 1, the 10 occurrences were of the same acoustic item each time. As a control measure, we also included an English Frequency factor, which has been shown to affect production of word-medial clusters (Munson, 2001).

3. Method

Materials were eight CVCCVC nonwords organized into four pairs. In one member of each pair, the medial cluster was frequent in English words, and in the other the cluster was infrequent (biphone transitional probabilities, calculated in Munson, 2001). The four pairs were: /fospəm foʃpəm/, /mæstəm mæfpəm/, /fæmpɪm fæmkɪm/, and /boktəm bopkəm/, with the first member containing the frequent cluster (High English) and the second the infrequent cluster (Low English). A token of each word was selected from a recording of an adult female speaker of American English. To manipulate Experiment Frequency, the item set was divided in half. Four of the items were heard 10 times (Experiment High) and the other four just once (Experiment Low). The two factors were crossed to yield four word types: High English-Experiment High, High English-Experiment Low, Low English-Experiment High, and Low English-Experiment Low. To avoid confounding English and Experimental Frequency, the items were distributed across two lists, with each word appearing as Experiment High in one list and Experiment Low in

the other. To reduce ordering effects, each of these lists was divided a second time: each word appeared in a different experimental block in a different list. This yielded four different participant lists.

Participants were 25 children ranging in age from 4;0 to 4;8, with a mean age of 4;5. Five children did not complete the study, four were reported by their mothers to have a personal or family history of speech, language or hearing abnormalities, and one was acquiring Spanish as their first language. The remaining 16 children (seven female) were included in the study. Four children received each list.

Procedure: children sat at a child-sized table and faced a computer screen with speakers. They were told that they would see pictures of some 'funny' animals and hear their names. Each child participated in two blocks. In each block, children were first familiarized with colored drawings of four novel animals paired with four of the eight nonwords. Pictures/nonwords were presented 10 times in the Experiment High condition and once in the Experiment Low condition, with four items (from two conditions) in each familiarization, and with token order randomized. Following familiarization, children participated in four test blocks – they were presented with each of the four pictures/nonwords in quasi-random order and were asked to repeat the word. Although no emphasis was placed on speed, children generally produced each nonword immediately. Stimulus presentation was controlled by SuperLab version 2.0 (Cedrus Corporation) running on a Macintosh G3 laptop (OS 9.2). The experimenter sat next to the child's table with the laptop next to her/him. At each juncture in the experiment, the experimenter explained what would happen next and only initiated that portion of the experiment when the child indicated readiness. Eight additional four-syllable words were included for a different experimental question (Goffman, Gerken, & Lucchesi, 2007).

3.1. Results and discussion

Two dependent measures were analyzed: production errors and production latency. With respect to production errors, each of the eight nonwords under consideration was transcribed and accuracy was totaled over the four consonants in each word. Each consonant was given a score of '0' for a correct production, '1' for an incorrect production, and '2' for a missing consonant. To assess reliability, eight (25%) of the words from each child were re-transcribed by a new transcriber. Only four (3%) of the words were re-transcribed with consonant differences that would have changed the error score, suggesting that the error data from the original transcript are reasonably reliable. Errors were summed over the four renditions of each nonword, and four consonants per rendition, for a maximum error score of 32. The error scores for the two words in each condition were then averaged so that each child contributed one production error data point for each of the four conditions. The mean number of errors made by each child was 4.38 (14%, see Figure 1a),

---

[1] In this paper, "talker variability" refers to differences in formant values (Peterson & Barney, 1952), fundamental frequency, articulatory dynamics (Ladefoged, 1980; Stevens, 1998), dialectal differences (Pierrehumbert, 2006), etc. "Affective variability" refers to changes in mean fundamental frequency and fundamental frequency range (Banse and Scherer, 1996; Scherer, 1986; Williams and Stevens, 1972), etc.
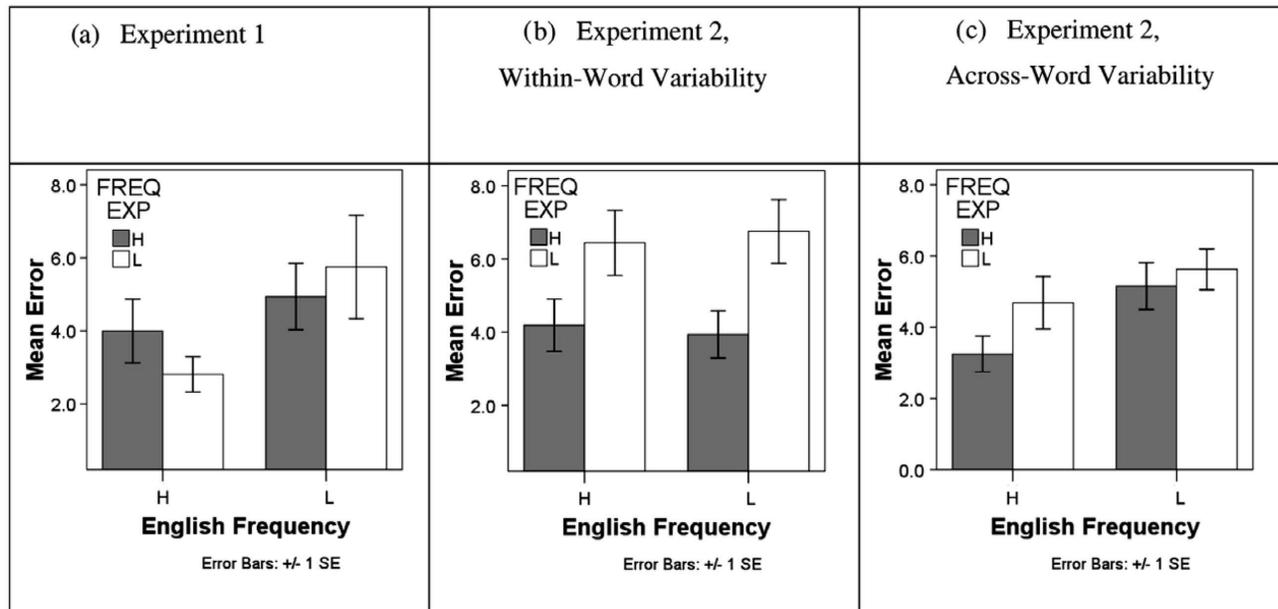
Figure 1. Graphs of production error means and standard errors in Experiments 1 and 2.

with a range of 0–12. A 2 English Frequency × 2 Experiment Frequency ANOVA was performed on the errors (see Figure 1a). There was a significant main effect of English Frequency, $F(1, 15) = 8.63$, $p = .01$, resulting from fewer errors on High English words, replicating the findings of Munson (2001). Neither the effect of Experiment Frequency nor the interaction was significant (both Fs < 1).

The second measure was the latency from the offset of the target word to the child's production, a measure thought to reflect production planning (Munson, 2001). Latencies were computed from digitized waveforms. Latencies 2 standard errors from the mean were omitted from

the analysis (8% of the latencies). They were also recalculated on 25% of each child's productions; 95% of the latencies were within 50 ms. of the original and 99% were within 100 ms. The average latency was 472.44 ms (see Figure 2a). A 2 English Frequency × 2 Experiment Frequency ANOVA was performed on the latencies for the four nonword types. There were no significant effects (all Fs < 1).

Although the error data from Experiment 1 replicated the effects of English phonotactic frequency that have been found by others, they showed no significant effect of the frequency with which a particular word appeared in the experiment. The lack of an effect of Experiment Frequency
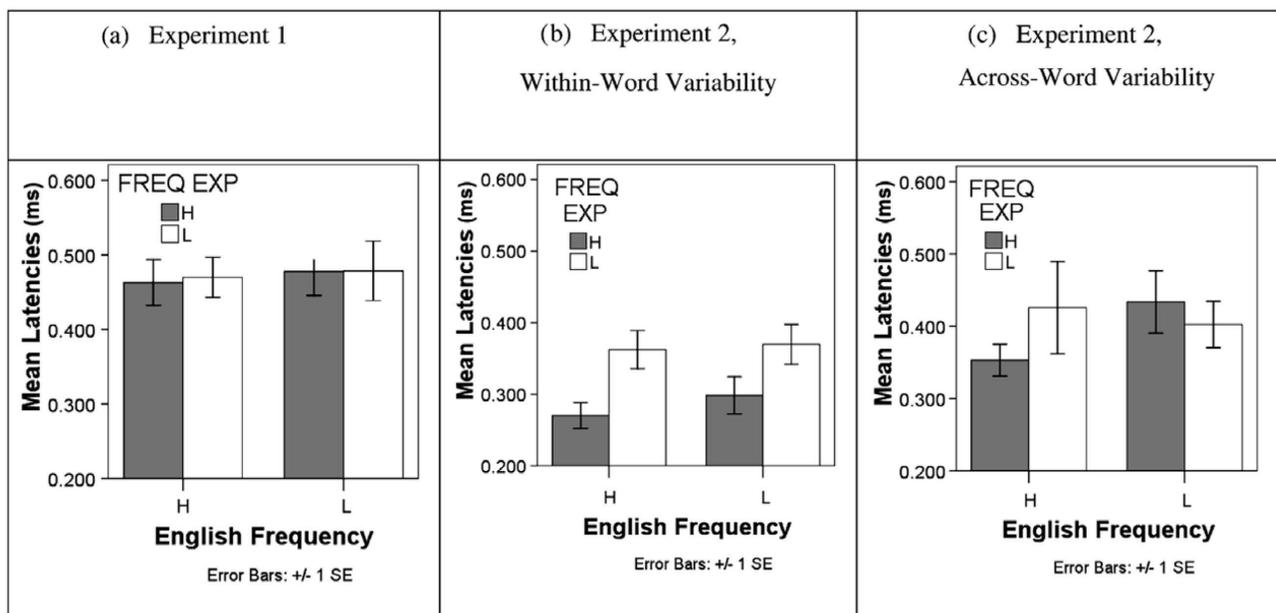

Figure 2. Graphs of production latency means and standard errors in Experiments 1 and 2.

in Experiment 1 might be taken to support the articulatory practice account of statistical frequency effects on production. However, the research on talker variability suggests that using the same acoustic stimulus 10 times in the two Experiment High conditions did little to encourage children to encode the words in a form abstract enough to be relevant to production (Houston, 2000; Singh, 2008). More generally, Experiment 1 makes the point that *frequency* is not a unified concept, but appears to be constrained in certain ways. Raw token frequency, for example, does not seem to be an appropriate concept for explaining why children might be more accurate when producing certain phonotactic sequences compared to others. Therefore, in Experiment 2, we introduced talker variability among the nonwords. In one condition (Within-Word Variability), children heard the Experiment High items produced by 10 different talkers. If talker variability allows children to form representations that better approximate a word's invariant properties, or the properties of a word that are relevant to production, then we should see children's productions significantly affected by this manipulation. In the second (control) condition (Across-Word Variability), children heard the items as produced by different talkers, but each word-token was acoustically identical. The purpose of this condition was to determine whether hearing multiple talkers was simply more interesting, and whether more interesting stimuli could lead to changes in production.

4. Experiment 2

Experiment 2 introduced talker variability to the Experiment Frequency factor. In line with the perceptual learning hypothesis, we predicted that this level of variability would affect production.

4.1. Method

Materials were identical to those used in Experiment 1. The Experimental Frequency factor differed, however, because the Experiment High nonwords were spoken by different talkers in the perception part of the experiment. Items in the Within-Word Variability Condition were either 10 tokens of each word spoken by 10 different talkers (Experiment High) or a single talker-token (Experiment Low). Items in the Across-Word Variability Condition were single tokens of the eight words heard 10 times (Experiment High) or once (Experiment Low), with each token produced by a different talker. Four different lists were created for both the Across-Word and Within-Word Conditions so that word-blocking was balanced. The target-word tokens used in the production test were identical to those used in Experiment 1 and were produced by a different talker than the talkers children heard during familiarization. In addition, the four-syllable words included in Experiment 1 were not used here.

Participants were 43 children ranging in age from 3;11 to 4;2, with a mean age of 4;1. Of those 43, four were removed

from the analysis because they did not complete the experiment, four were inaudible, two were reported to have a speech, language, or hearing delay, and one was removed due to an experimenter error. For the remaining 32 children (16 female), half were assigned to the Across-Word condition and half to the Within-Word condition.

Procedure: the procedure was identical to that of Experiment 1.

4.2. Results and discussion

As in Experiment 1, production errors and production latencies were analyzed. Each of the eight nonwords under consideration was transcribed, and accuracy was totaled over the four consonants in each word. For participants in the Within-Word Condition, eight (25%) of the words from each child were re-transcribed. With respect to discrepancies, none of the consonant differences would have changed the error score, suggesting that the original transcription was reliable. The mean number of errors made by each child was 5.33 (17%, see Figure 1b and c), with a range of 0–11, similar to what was found in Experiment 1. For participants in the Across-Word Condition, all of the words were re-transcribed by a new transcriber, then the two transcribers resolved discrepancies in a joint meeting.[2] The mean number of errors made by each child was 4.68 (15%, see Figure 1c), with a range of 0–11, similar to the errors found in Experiment 1 and in the Across-Word Condition.

A 2 English Frequency × 2 Experiment Frequency × 2 Variability ANOVA was performed on the errors for the eight nonword types (see Figure 1b). Consistent with the perceptual learning hypothesis, there was a main effect of Experiment Frequency, $F(1, 30) = 9.78$, $p < .01$, although there were no effects of English Frequency or Variability, and no significant interactions. The absence of an effect of English Frequency may have resulted because the local effects of frequency that were manipulated in the experiment were more potent at the time of production than the background frequency statistics of the child's lexicon. The absence of an effect of Variability suggests that Within-Word and Across-Word variability were equally effective in facilitating children's production accuracy, although the graphs of the two conditions in Figure 1 suggest that the Within-Word Variability Condition was the primary contributor to the effect of Experiment Frequency.

As in Experiment 1, production latencies were computed from the waveforms. Latencies that were 2 standard errors from the mean were omitted from the analysis (6% of Within-Word latencies, 2% of Across-Word latencies). For both the Within-Word and Across-Word Conditions, latencies were recalculated on 25% of each child's productions. For the Within-Word Condition, 93% of the latencies were within 50 ms. of the original and 95% were within 100 ms. The average latency was 330.97 ms. For the Across-Word Condition, 85% of the latencies were within 50 ms. of each other and 95% were within 100 ms. The average latency was 403.6 ms (see Figure 1b and c).

---

[2] The Within-Word and Across-Word Variability Conditions were run with separate participants and at separate times. In the intermediary, a change in research staff necessitated the change in reliability analyses.

A 2 English Frequency × 2 Experiment Frequency × 2 Variability ANOVA was performed on the latencies. There were main effects of English Frequency, $F(1, 30) = 11.24$, $p < .01$, resulting from faster reaction times for words containing high frequency clusters, and an effect of Experiment Frequency, $F(1, 30) = 17.80$, $p < .01$, resulting from faster reaction times for items that children heard 10 times during familiarization. More importantly, there was not a three-way interaction ($F < 1$), but there was a significant two-way Experiment Frequency × Variability interaction, $F(1, 30) = 10.38$, $p < .01$, and a trend towards an English Frequency × Variability interaction, $F(1, 30) = 3.16$, $p = .086$. Considering the simple effects for the Experiment Frequency × Variability result, there was a significant effect of Experiment Frequency in the Within-Words Condition, $F(1, 15) = 23.80$, $p < .01$, but not in the Across-Words Condition, $F(1, 15) = 0.59$, $p = .45$, suggesting that children were faster as a result of hearing a word produced by multiple talkers. For the English Frequency × Variability trend, there was *not* a significant effect of English Frequency in the Within-Words Condition, $F(1, 15) = .96$, $p = .34$, but there was a significant effect of English Frequency in the Across-Words Condition, $F(1, 15) = 18.59$, $p < .01$. In sum the two-way interactions in the production latency analysis strengthen the conclusions suggested by the errors analysis. First, we see clear support for the perceptual learning hypothesis in children's production latencies, as children were consistently faster to produce words that they heard 10 times during the familiarization. However, children's speed was more influenced by the within-word variability present in 10 talker-tokens than by the across-word talker variability. This follows nicely from the benefits of variability seen in studies of adult (Goldinger, 1996; Johnson, 1997) and infant word recognition (Houston, 2000; Singh, 2008). In the latter literature, infants showed more robust word recognition following familiarization with multiple talkers or multiple affective qualities. In the present experiment, children were faster and more accurate to produce novel words when they heard them spoken by different talkers, but their production latencies were most sensitive to variable productions of the same word. Finally, the English Frequency × Variability analysis and the apparent absence of an English Frequency effect in the within-word variability condition suggests that within-word variability may enhance production planning in the short-term more than language-wide frequencies do.

5. General discussion

The results of these experiments suggest that speech production is, at least in part, dependent on *perceptual learning*. These studies do not rule out some role for articulatory practice, but it appears that perceptual learning alone is sufficient to change production speed and accuracy. Future studies may consider whether and how articulatory practice and perceptual learning combine to influence production.

The data from Experiments 1 and 2 also make two important points about the nature of perceptual learning.

First, Experiment 2 indicates that statistical frequency in the input can affect children's language production directly. Previous studies of adult L2 acquisition showed a similar result (Bradlow et al., 1997, 1999; Wang et al., 2003), but to our knowledge, no finding of this sort exists in the child language literature. We take these results as evidence that ambient linguistic input can be encoded in a sufficiently abstract form for use in language production.

Second, the data reveal that perceptual learning is more than just statistics; it is supported by phonetic variation. In Experiment 1, in which the Experimental Frequency manipulation was limited to raw token frequency, no clear effects of perceptual learning were found. In contrast, the robust effects of Experiment Frequency in Experiment 2 converge with results from infant studies showing that talker- or affective-based variability is an important component of perceptual processing (e.g., Houston, 2000; Singh, 2008). The results from Experiment 2 also complement perceptual learning studies in which variability cued learners to the relevant dimensions of an acoustic category (e.g., Holt & Lotto, 2006). More broadly, machine learning research stresses the importance of a variable input—a variety of data points are necessary for a learner to be able to establish a category and its boundaries. Talker variability may be just one type of variability that facilitates learning, establishing a robust representation of a word. However, further research is needed to understand whether our participants were engaged in holistic word learning, that is, whether they learned about producing whole words by hearing those words during familiarization, or whether they learned something about phonotactics, as is suggested by the improvements to production speed and accuracy for matched high and low probability clusters.

Acknowledgments

References

Banse and Scherer, 1996 ▶ R. Banse and K. R. Scherer, Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology* **70** (3) (1996), pp. 614–636.

Bradlow et al., 1999 ▶ A. R. Bradlow, R. Akahane-Yamada, D. B. Pisoni, and Y. Tohkura, Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production, *Perception and Psychophysics* **61** (5) (1999), pp. 977–985.

Bradlow et al., 1997 ▶ A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada, and Y. Tohkura, Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production, *Journal of the Acoustical Society of America* **101** (4) (1997), pp. 2299–2310.

Chambers et al., 2003 ▶ K. E. Chambers, K. H. Onishi, and C. L. Fisher, Infants learn phonotactic regularities from brief auditory experience, *Cognition* **87** (2003), pp. B69–B77.

Edwards et al., 2004 ▶ J. Edwards, M. E. Beckman, and B. Munson, The interaction between vocabulary size and phonotactic probability on children's production accuracy and fluency in nonword repetition, *Journal of Speech, Language, and Hearing Research* **47** (2004), pp. 421–436.

Goffman et al., 2007 ▶ L. Goffman, L. A. Gerken, and J. Lucchesi, Relations between segmental and motor variability in prosodically complex nonword sequences, *Journal of Speech, Language, and Hearing Research* **50** (2007), pp. 444–458.

Goldinger, 1996 ▶ S. D. Goldinger, Words and voices: Episodic traces in spoken word identification and recognition memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22** (1996), pp. 1166–1183.

Goldinger, 1998 ▶ S. D. Goldinger, Echoes of echoes? An episodic theory of lexical access, *Psychological Review* **105** (2) (1996), pp. 251–279.

Guenther, 2006 ▶ F. H. Guenther, Cortical interactions underlying the production of speech sounds, *Journal of Communication Disorders* **39** (5) (2006), pp. 350–365.

Holt and Lotto, 2006 ▶ L. L. Holt and A. J. Lotto, Cue weighting in auditory categorization: Implications for first and second language acquisition, *Journal of the Acoustical Society of America* **119** (5) (2006), pp. 3059–3071.

Houston, 2000 ▶ D. M. Houston. *The role of talker variability in infant word representations*. Johns Hopkins University: Unpublished doctoral dissertation, 2000.

Johnson, 1997 ▶ K. Johnson, Speech perception and speaker normalization. In: K. Johnson and J. Mullenix, Editors, *Talker Variability in Speech Processing*, Academic Press, San Diego, CA (1997).

Johnson and Jusczyk, 2001 ▶ E. K. Johnson and P. W. Jusczyk, Word segmentation by 8-month-olds: When speech cues count more than statistics, *Journal of Memory and Language* **44** (4) (2001), pp. 548–567.

Jusczyk et al., 1993 ▶ P. W. Jusczyk, A. D. Friederici, J. M. Wessels, V. Y. Svenkerud, and A. M. Jusczyk, Infants' sensitivity to the sound patterns of native language words, *Journal of Memory and Language* **32** (3) (1993), pp. 402–420.

Jusczyk et al., 1994 ▶ P. W. Jusczyk, P. A. Luce, and J. Charles-Luce, Infants' sensitivity to phonotactic patterns in the native language, *Journal of Memory and Language* **33** (1994), pp. 630–645.

Ladefoged, 1980 ▶ P. Ladefoged, What are linguistic sounds made of?, *Language* **56** (1980), pp. 485–502.

MacKay, 1989 ▶ D. G. MacKay, The organization of perception and action: A theory of language and other cognitive skills, Springer, New York, NY (1989).

Maye et al., 2002 ▶ J. Maye, J. Werker, and L. A. Gerken, Infant sensitivity to distributional information can affect phonetic discrimination, *Cognition* **82** (3) (2002), pp. B101–B111.

Munson, 2001 ▶ B. Munson, Phonological pattern frequency and speech production in adults and children, *Journal of Speech, Language and Hearing Research* **44** (2001), pp. 778–792.

Peterson and Barney, 1952 ▶ G. E. Peterson and H. L. Barney, Control methods used in a study of the vowels, *Journal of the Acoustical Society of America* **24** (1952), pp. 309–328.

Pierrehumbert, 2006 ▶ J. B. Pierrehumbert, The next toolkit, *Journal of Phonetics* **34** (2006), pp. 516–530.

Saffran et al., 1996 ▶ J. R. Saffran, R. Aslin, and E. Newport, Statistical learning by 8-month-old infants, *Science* (1996).

Saffran and Thiessen, 2003 ▶ J. R. Saffran and E. D. Thiessen, Pattern induction by infant language learners, *Developmental Psychology* **39** (2003), pp. 484–494.

Scherer, 1986 ▶ K. Scherer, Vocal affect expression: A review and a model for future research, *Psychological Bulletin* **9** (1986), pp. 143–165.

Singh, 2008 ▶ L. Singh, Influences of high and low variability on infant word recognition, *Cognition* **106** (2) (2008), pp. 833–870.

Stevens, 1998 ▶ K. N. Stevens, Acoustic phonetics, MIT Press, Cambridge, MA (1998).

Storkel, 2001 ▶ H. L. Storkel, Learning new words: Phonotactic probability in language development, *Journal of Speech, Language, and Hearing Research* **44** (2001), pp. 1321–1337.

Storkel, 2003 ▶ H. L. Storkel, Learning new words II: Phonotactic probability in verb learning, *Journal of Speech, Language, and Hearing Research* **46** (2003), pp. 1312–1323.

Wang et al., 2003 ▶ Y. Wang, A. Jongman, and J. A. Sereno, Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training, *Journal of the Acoustical Society of America* **113** (2) (2003), pp. 1033–1043.

Williams and Stevens, 1972 ▶ C. E. Williams and K. N. Stevens, Emotions and speech: Some acoustical correlates, *Journal of the Acoustical Society of America* **52** (4B) (1972), pp. 1238–1250.

Zamuner et al., 2004 ▶ T. Zamuner, L. A. Gerken, and M. Hammond, Phonotactic probabilities in young children's speech productions, *Journal of Child Language* **31** (2004), pp. 515–536.

Zamuner et al., 2005 ▶ T. Zamuner, L. A. Gerken, and M. Hammond, The acquisition of phonology based on input: A closer look at the relation of cross-linguistic and child language data, *Lingua* **115** (10) (2005), pp. 1329–1474.