

Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English

Eon-Suk Ko

University at Buffalo, State University of New York, Buffalo, New York 14260
eonsukko@buffalo.edu

Melanie Soderstrom

University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
m_soderstrom@umanitoba.ca

James Morgan

Brown University, Providence, Rhode Island 02912
james_morgan@brown.edu

Abstract: 8- and 14-month-old infants' perceptual sensitivity to vowel duration conditioned by post-vocalic consonantal voicing was examined. Half the infants heard CVC stimuli with short vowels, and half heard stimuli with long vowels. In both groups, stimuli with voiced and voiceless final consonants were compared. Older infants showed significant sensitivity to mismatching vowel duration and consonant voicing in the short condition but not the long condition; younger infants were not sensitive to such mismatching in either condition. The results suggest that infants' sensitivity to extrinsic vowel duration begins to develop between 8 and 14 months.

© 2009 Acoustical Society of America

PACS numbers: 43.71.Ft, 43.70.Mn, 43.71.Es [JH]

Date Received: July 14, 2009 **Date Accepted:** August 31, 2009

1. Introduction

The development of phonological category knowledge involves two important components. Infants' perceptual systems must be tuned to the phoneme boundaries that exist in their native language, and they must be sensitive to systematic subphonemic variations in which the location of phoneme boundaries is influenced by variations along other acoustic dimensions. One example of the latter type is the relationship between the vowel length and the perception of a coda consonant as voiced or voiceless. In this study, we examine infants' sensitivity to the property of consonant voicing in the context of short and long vowel durations.

Young infants demonstrate sensitivity to within-category subphonemic distinctions in voice onset time (Miller and Eimas, 1996). Similarly, infants are sensitive to the allophonic variation of aspiration in isolation at 2 months (Hohne and Jusczyk, 1994) and are able to use allophonic information as a cue to identify familiarized target words in fluent speech by the age of 10.5 months (Jusczyk *et al.*, 1999). Infants are therefore able to detect at least some subphonemic variations, when they do not affect the perception of phoneme boundaries. What is unknown is how these sensitivities influence infants' phonological representations, when those variations are relevant to native-language-like perception of phoneme distinctions.

The present article investigates infants' development of perceptual sensitivity to subsegmental phonotactics, focusing on variation in vowel duration conditioned by the voicing of the following consonant. Vowels are realized with longer duration before a voiced than a voiceless consonant in English, e.g., [pɪk] vs [pɪ:g] (House and Fairbanks, 1953). This effect will be referred to as "vowel length effect" (VLE). The duration of a pre-consonantal vowel thus serves as a source of information about the voicing of the following consonant. In addition to the VLE, earlier research has examined aspects of other cues for the post-vocalic voicing such as F1

offset frequency (Fischer and Ohde, 1990), intensity decay time, and the presence or absence of a “voice bar” during the closure interval (Hillenbrand *et al.*, 1984). The focus of our investigation was on the development of sensitivity to VLE-induced phonotactics.

Adult English listeners weight vocalic duration strongly in their perceptual decisions about the voicing of final stops, especially when no release burst is present (Denes, 1955) or the stimuli are synthetic (Raphael, 1972). However, 5–10 year old children and adults tested with stimuli based on natural utterances attend largely to dynamic signal components such as the F1-offset transitions rather than the vocalic duration (Morrongiello *et al.*, 1984, although see Hillenbrand *et al.*, 1984). Eilers (1977) suggested that infants at around 2 months of age use vowel duration as a supplementary cue for discriminating final consonantal voicing. Eilers *et al.* (1984) similarly found that infants (5–11 months) have the ability to discriminate vowel duration differences but their performance was much poorer than that of adults. Both studies examined instances of lengthening but not shortening. Lengthening differs from shortening, in that higher-level prosodic effects can also cause vowels to be lengthened, for example when words are focally or emphatically stressed, or occur at the ends of phonological or intonational phrases. These other factors may well complicate infants’ reactions to vowel lengthening.

Recently, Dietrich *et al.* (2007) found that Dutch and English learning 18-month-olds treat vowel duration differently in a word learning task. In Dutch, vowel duration is an important cue for differentiating the low vowels [ɑ] and [a:], whereas in English, it is only a secondary cue to distinguish a tense from a lax vowel. Their results indicate that these properties are reflected in infants’ perceptual sensitivity: Dutch learners interpret vowel duration as lexically contrastive, whereas English learners do not. One might therefore predict that 18-month-old English learners do not discriminate vowel duration differences. However, a subsequent study by Mugitani *et al.* (2009) found that 18-month-old English learners discriminate vowel duration differences if the task does not require linking objects with words. They also found that, in Japanese, where vowel length is phonemic, younger infants (10-month-olds) discriminate vowel duration differences like English 18-month-olds, while 18-month-olds show an asymmetric pattern of discrimination, responding to shortening, but not lengthening, of the vowel.

What do these findings suggest about infants’ knowledge of the phonotactic patterns characterized by VLE? As noted, subphonemic variation in vowel duration can serve as a cue to post-vocalic voicing. Is infants’ sensitivity to this pattern an innate characteristic of the perceptual system, or does it develop through exposure to the distributional characteristics of the language? Cross-linguistic comparisons suggest that speakers of languages without the VLE do not rely on vowel duration as a cue to voicing as much as do speakers of languages with the VLE. The use of the VLE as a perceptual cue may be learned through the experience with a native language (Crowther and Mann, 1992). Such language-specific patterns in perceptual weighting strategies predict that infants learning American English must acquire their sensitivity to the VLE at some point. Given that infants’ speech perception is largely native-like by around 12 months (Werker and Tees, 1984), a year’s exposure to English may have provided enough information for infants to develop their perceptual sensitivity to the VLE. However, there is relatively little work on their perception of coda consonants.

The present study investigated the development of 8- to 14-month-olds’ perceptual sensitivity to the VLE. These ages roughly correspond to the beginning and end of the period of attunement toward native-like phoneme perception. Infants’ first words also emerge toward the end of this period, giving us an opportunity to relate the results of their perceptual sensitivity to the patterns of the VLE in their early speech production. Recent findings (Ko, 2007) suggest that infants’ learning of the VLE may have already begun to develop by the onset of their speech production. We hypothesized that infants by 14 months may have begun to develop their sensitivity to the VLE. We presented half the infants with CVC syllables containing a long vowel followed by a voiced (matched) or a voiceless (mismatched) consonant, and the other half with syllables containing a short vowel followed by a voiced or a voiceless consonant. If infants detect the relationship between vowel duration and coda voicing, they should discriminate matched from mismatched trials.

Table 1. Mean naturalness scores for mismatch tokens.

Token	bæ:k	bæg	kɔ:p	kɔb	pɪ:k	pɪg
Mean naturalness score ($n=3$)	3.3	3.2	3.6	3.5	3.2	3.7

2. Method

Subjects. Seventy infants were tested; thirty-three 8-month-olds, and thirty-seven 14-month-olds. Four participants in the 14-month-old group were excluded from analysis because of fussiness ($n=2$) or lack of interest in the study ($n=2$). One participant in each of the 8- and 14-month-old groups was excluded due to experimenter error. This left thirty-two 8-month-olds (16 boys and 16 girls, mean age = 257 days, age range = 241–290 days) and thirty-two 14-month-olds (19 males and 13 females, mean age = 432 days, age range = 411–451 days). Half the infants heard words with a long vowel, followed by either a voiced (*matched long*; [pɪ:g, kɔ:b, bæ:g]) or a voiceless consonant (*mismatched long*; [pɪ:k, kɔ:p, bæ:k]). The other half heard *matched short* ([pɪk, kɔp, bæ:k]) and *mismatched short* ([pɪg, kɔb, bæ:g]) stimuli.

Stimuli. The stimuli were constructed from three minimal pairs ending in a voiced/voiceless plosive, *bag/back*, *cup/cup*, and *pig/pick*. A female native speaker of American English spoke the base words multiple times with infant-directed prosody, using a strong coda release. This ensured that perceptual cues for voice distinction associated with the release of a plosive were available in the stimuli, eliminating the possibility of cues other than vowel duration interfering as a confounding factor in the perception of the VLE-induced patterns.

Six exemplars of each word were chosen as the base tokens to produce the final stimuli by manipulating the duration of the vowel. The stimuli underwent lengthening/shortening of the vowel using the PSOLA resynthesis method available in PRAAT (Boersma and Weenink, 2007). Mismatched stimuli were constructed by lengthening or shortening the nucleus vowel of the base token, and matched stimuli were generated by lengthening or shortening the mismatched stimuli back to the original vowel duration. We generated the matched tokens through manipulation rather than using the natural base tokens to prevent any confounding effects of infants' perception of or preferences for natural vs manipulated stimuli. The resulting stimuli contain all the cues for the post-vocalic voice distinction such as pitch and formant transitions except for the vowel duration. The degrees of lengthening and shortening were 160% and 50% of the nucleus vowel in the base token.

The resulting 36 mismatched tokens (6 exemplars \times 6 words) were rated for naturalness by ten adult subjects. The purpose of this testing was to ensure that the lengthened and shortened mismatched stimuli maintain about the same level of naturalness. The stimuli were presented in randomized order using PRAAT, and subjects scored the naturalness of each token from the scale of 1 (least natural) to 5 (most natural). Based on the results of the naturalness ratings, we selected 18 final tokens of mismatched stimuli (3 exemplars \times 6 words) that yielded balanced naturalness ratings between lengthened and shortened tokens (see Table 1). The average vowel duration in the base tokens for these final tokens are reported in Table 2. Based on these 18 mismatched stimuli, we constructed 18 matched stimuli by manipulating the vowel duration back to the original base tokens.

Procedure. Testing was performed in a sound-attenuated room using the Headturn Preference Procedure. The testing booth consisted of a three-walled enclosure made of white

Table 2. Mean duration of the vowel in base tokens.

Token	bæ:g	bæk	kɔ:b	kɔp	pɪ:g	pɪk
Mean duration in ms ($n=3$)	297.4	119.5	166.8	95.6	215.2	110.4

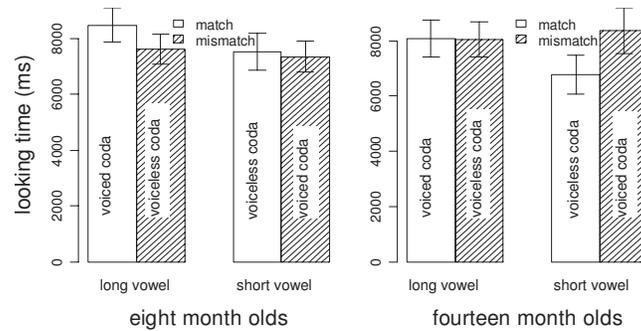


Fig. 1. Infants' preferences for VLE-matching vs mismatching word tokens.

pegboard panels, with a light mounted at the center of each panel wall. Caregivers sat with their infant on their lap and wore aviator headphones which played masking music to avoid biasing the infant's behavior. The order of trial presentation was randomized on-line by the experimental software.

Each trial began with the front light blinking to attract the infants' attention. When the infant looked at the center light, one of the two side lights began to flash. When the infant looked toward that light, the stimuli for that trial played from a speaker behind the light. Infants were first presented with two practice trials containing repetitions of three tokens of *book* and *dog*. They were immediately followed by a testing session of two randomized blocks of six trials containing the matched and mismatched versions of the 6 test words (12 test trials). Each trial consisted of random repetitions of the three exemplars of a particular word. Thus the "long" group heard tokens of [bæ:g], [bæ:k], [kɹ:b], [kɹ:p], [pɪ:g], and [pɪ:k] on successive trials, and the "short" group heard the short counterpart of each of these stimuli. Since word tokens varied considerably in length, pause durations between tokens for each trial were chosen in order to maintain a consistent interval between the onsets of each stimulus at 1200 ms. Therefore, infants heard similar rates of token presentation across trials and conditions. The dependent variable was the average amount of time each infant listened to matched vs the mismatched stimuli, based on their looking behavior.

3. Results

Mean looking times for stimuli with long and short vowels before voiced and voiceless coda consonants are shown in Fig. 1 for each of the two age groups tested. An analysis of variance (ANOVA) with two between-subjects factors, age and duration (short/long), and one within-subjects factor, matching, found significant interactions between age and matching, $F(1, 60) = 5.46, p < 0.05$, and between duration and matching, $F(1, 60) = 4.36, p < 0.05$ (see Fig. 1). Individual ANOVAs for each between-subjects condition found a significant interaction between matching and age, $F(1, 30) = 5.34, p < 0.05$, and a marginal main effect of matching, $F(1, 30) = 3.47, p = 0.072$, in the short condition, but no main effect or interactions in the long condition. In the 14-month-old age group, a marginal interaction was found for matching and duration, $F(1, 30) = 4.12, p = 0.051$, with a marginal main effect of matching, $F(1, 30) = 3.79, p = 0.061$. No significant effects were found with the 8-month-olds. Overall, these effects and interactions reflect a significant preference for the mismatched stimuli (mean listening time = 8.4 s) over the matched stimuli (mean listening time = 6.9 s) in the short condition for the older infants only, $t(15) = 3.20, p < 0.01$.

In sum, 14-month-olds showed a significant sensitivity to the mismatching of vowel duration and consonant voicing in the short, but not the long condition. Eight-month-olds did not show sensitivity with either short or long vowels.

4. Discussion

Our data suggest that sensitivity to the VLE develops over the course of the second half of the first year of life, consistent with the view that it is acquired through experience with phonotactic patterns in the native language. This is convergent with findings that speakers of languages without the VLE use vowel duration less than speakers of languages with the VLE (Crowther and Mann, 1992). It is also consistent with the recent finding that language-specific phonology influences the development of infants' speech perception (Mugitani *et al.*, 2009).

At first blush, our findings appear to contradict Dietrich *et al.* (2007), in which English-learning 18-month-olds failed to link two novel objects with the two stimuli differing only in vowel duration. However, there is good reason to suspect that older infants are less likely to discriminate auditory patterns in a word-learning context than in a pure preference or discrimination task (Stager and Werker, 1997). Therefore, it may be that 18-month-old English learners retain perceptual sensitivity to the VLE, as suggested in Mugitani *et al.* (2009), but fail to demonstrate this ability in a word-learning task: the oddness of a mismatch between vowel duration and coda voicing may not be regarded as encoding a lexical distinction.

The asymmetry between short and long vowels in our study may well be a consequence of infants' familiarity with vowel lengthening effects such as phrase-final lengthening and vowel elongation in infant-directed speech. Vowels are lengthened due to a variety of causes, and thus long vowels appear in variable contexts in the input. Therefore, infants may treat shortening as a more relevant cue for the phoneme boundary than lengthening or treat lengthening as more acceptable than shortening. This is consistent with our observation that 14-month-olds discriminated matched and mismatched exemplars containing short vowels, but not exemplars containing long vowels. Similar findings of such asymmetry are reported in other studies. For example, Hogan and Rozsypal (1980), testing the effects of vowel modulation on adults' judgment of voice distinction for post-vocalic consonants, reported findings of a pilot study in which recognition of the stimuli ending with a voiceless consonant remained unaffected by lengthening of the vowel. More recently, Japanese 18-month-old infants (Mugitani *et al.*, 2009) and Dutch 21-month-old toddlers (van der Feest and Swingley, 2008) have been reported to show asymmetric discrimination patterns to the vowel duration change. These findings suggest different processing of lengthening and shortening in infants as well as adults.

Given the stimuli we used, it is possible that 14-month-olds perceived short vowels preceding voiced consonants as aberrant pronunciations of familiar words, rather than as violations of more general phonotactic patterns. We plan to tease these possibilities apart in a follow-up study using nonce word stimuli.

Our results indicate that the perceptual system of 14-month-olds, who are at the beginning stages of word production, is already sensitive to the VLE, at least in some contexts. This suggests that the emergence of the VLE in children's early speech (Ko, 2007) may reflect children's knowledge of English phonotactics in the perceptual domain. The current study thus provides some concrete data to corroborate the idea that the development of speech production is preceded by the development of perceptual sensitivity.

5. Conclusion

The current study examined infants' development of perceptual sensitivity to the VLE. Our findings suggest that infants' sensitivity to the phonotactic patterns conditioned by the VLE begin to develop between 8 and 14 months. Infants may begin to use vowel duration as a cue to voicing at least as early as 14 months. Our results also point to an asymmetry supported by a growing body of research indicating that lengthening and shortening effects are treated differently in the speech perception.

Acknowledgments

This study was supported by NIH Grant No. R01 HD23005 to J.L.M. We thank Lori Rolfe, Elena Tenenbaum, Erin Conwell, Jae Yung Song, Amanda Seidl, Alex Cristià, and the participants of the experiments for their help in completing this study.

References and links

- Boersma, P., and Weenink, D. (2007). PRAAT: Doing Phonetics by Computer Version 4.6.38, from <http://www.praat.org/> (Last viewed November, 2009).
- Crowther, C. S., and Mann, V. A. (1992). "Native language factors affecting use of vocalic cues to final consonant voicing in English," *J. Acoust. Soc. Am.* **92**, 711–722.
- Denes, P. (1955). "Effect of duration on the perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.
- Dietrich, C., Swingle, D., and Werker, J. F. (2007). "Native language governs interpretation of salient speech sound differences at 18 months," *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16027–16031.
- Eilers, R. (1977). "Context-sensitive perception of naturally produced stop and fricative consonants by infants," *J. Acoust. Soc. Am.* **61**, 1321–1336.
- Eilers, R., Bull, D., Oller, K., and Lewis, D. (1984). "The discrimination of vowel duration by infants," *J. Acoust. Soc. Am.* **75**, 1213–1218.
- Fischer, R. M., and Ohde, R. N. (1990). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," *J. Acoust. Soc. Am.* **88**, 1250–1259.
- Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (1984). "Perception of the voiced-voiceless contrast in syllable-final stops," *J. Acoust. Soc. Am.* **76**, 18–26.
- Hogan, J., and Rozsypal, A. (1980). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," *J. Acoust. Soc. Am.* **67**, 1764–1771.
- Hohne, E., and Jusczyk, P. (1994). "Two-month-old infants' sensitivity to allophonic differences," *Percept. Psychophys.* **56**, 613–623.
- House, A., and Fairbanks, G. (1953). "The influence of consonantal environment upon the secondary acoustical characteristics of vowels," *J. Acoust. Soc. Am.* **25**, 105–113.
- Jusczyk, P., Hohne, E., and Bauman, A. (1999). "Infants' sensitivity to allophonic cues for word segmentation," *Percept. Psychophys.* **61**, 1465–1476.
- Ko, E. (2007). "Acquisition of vowel duration in children speaking American English," in *Proceedings of Interspeech 2007*, pp. 1881–1884.
- Miller, J., and Eimas, P. (1996). "Internal structure of voicing categories in early infancy," *Percept. Psychophys.* **58**, 1157–1167.
- Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. (1984). "Trading relations in the perception of speech by 5-year-old children," *J. Exp. Child Psychol.* **37**, 231–250.
- Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J., and Amano, S. (2009). "Perception of vowel Length by Japanese- and English-learning infants," *Dev. Psychol.* **45**, 236–247.
- Raphael, L. J. (1972). "Preceding vowel duration as a cue to the voicing characteristics of word-final consonants in English," *J. Acoust. Soc. Am.* **51**, 1296–1303.
- Stager, C. L., and Werker, J. F. (1997). "Infants listen for more phonetic detail in speech perception than in word learning tasks," *Nature (London)* **388**, 381–382.
- van der Feest, S., and Swingle, D. (2008). "A crosslinguistic study of vowel duration in 21-month-olds' early lexical representations," paper presented at the 16th International Conference on Infant Studies, Vancouver, Canada.
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.