

# Voice-onset time in infant-directed speech over the first year and a half

Evamarie Burnham Jessica L. Gamache Tonya Bergeson Laura Dilley

Citation: **19**, 060094 (2013); doi: 10.1121/1.4800072

View online: <http://dx.doi.org/10.1121/1.4800072>

View Table of Contents: <http://asa.scitation.org/toc/pma/19/1>

Published by the [Acoustical Society of America](#)

---

---



**ICA 2013 Montreal  
Montreal, Canada  
2 - 7 June 2013**

**Speech Communication**

**Session 2aSC: Linking Perception and Production (Poster Session)**

**2aSC50. Voice-onset time in infant-directed speech over the first year and a half**

Evamarie Burnham, Jessica L. Gamache, Tonya Bergeson and Laura Dilley\*

\*Corresponding author's address: Dept. of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI 48824, [ldilley@msu.edu](mailto:ldilley@msu.edu)

Previous research in small-N studies has indicated conflicting findings regarding whether mothers modify voice-onset time (VOT) of word-initial stop consonants in speech to infants compared to speech to adults, as well as the nature of any such modification. In a large-scale study, VOT was measured for voiced and voiceless stop consonants in speech of 48 mothers of infants in one of four cross-sectional age groups (0;3, 0;9, 1;1, 1;8) when they read a phonetically-controlled storybook to their infant (ID speech) or an adult (AD speech). VOT measurements showed enhanced clarity (i.e., longer VOTs) in ID speech compared with AD speech for voiceless stop consonants only. An effect of infant gender was also found, showing that enhanced clarity was only produced by mothers of female infants ( $N = 19$ ). Infant age was not found to be a significant factor in VOT production. The results have implications for understanding the nature of linguistic development in young children, specifically by elucidating factors apparently related to phonetic modification for clarity, including speech style and gender. Work supported by NIH-NIDCD grant R01DC008581.

Published by the Acoustical Society of America through the American Institute of Physics

## INTRODUCTION

A growing body of research ties clear speech input during infancy to improved speech development (e.g., Liu, Kuhl, and Tsao, 2003; Tsao, Liu, and Kuhl, 2004; Kuhl, Conboy, Padden, Nelson, and Pruitt, 2005). Studies of infant-directed (ID) speech have identified a number of prosodic and segmental features that mothers produce in a clearer or more exaggerated manner in ID speech than adult-directed (AD) speech. For prosodic features, mothers exhibit increased pitch range, slower speech rate, and longer pauses in ID speech as compared to AD speech (Snow, 1977; Fernald and Simon, 1984; Papoušek, Papoušek and Bornstein, 1985; Fernald, 1992; Bergeson and Trehub, 2002; Bergeson, Miller, and McCune, 2006). For segmental features, mothers appear to more carefully articulate segments in ID speech as compared to AD speech, including by increasing formant differences between point vowels (Kuhl, Andruski, Chistovich, Chistovich, Kozhevnikova, Ryskina, *et al.*, 1997; Burnham, Kitamura, and Vollmer-Conna, 2002; Liu *et al.*, 2003), producing greater differentiation of voice onset times (VOTs) contrasting voiced vs. voiceless stop consonants (Malsheen, 1980; Englund, 2005; Sundberg, 2001), and using less phonological reduction (Bernstein Ratner, 1984).

However, other studies have found that AD speech can show segmental characteristics that either are as clear as, or clearer than, those of ID speech. For example, while Bernstein Ratner (1984) found less phonological reduction in ID speech as compared to AD speech, Shockey and Bond (1980) found that phonological reduction was more likely in child-directed than AD speech. Similarly, Bernstein Ratner and Luberoff (1984) found that word final consonants were more likely to be deleted or glottalized in child-directed speech than in AD speech. For VOTs, while Malsheen (1980), Sundberg (1999), and Englund (2005) found that VOTs were more differentiated in ID speech than AD speech, Baran, Laufer, and Daniloff (1977), Sundberg (2001) and Synnestvedt (2010) found that VOTs were less differentiated in ID speech than AD speech.

There are many reasons why these disparities may exist between results, chief among them small numbers of participants, differences in gender of the infants, and variability in age of the children. Small numbers of participants can lead to spurious findings resulting from underrepresentation; many of the studies cited above involved less than 10 participants across a wide range of ages (Malsheen, 1980; Sundberg and Lacerda, 1999; Englund, 2005; Shockey and Bond, 1980; Bernstein Ratner, 1984; Baran *et al.*, 1977). The gender of the infant has also been identified as a factor that may affect mothers' clarity in ID speech; Dilley, Millett, McAuley, and Bergeson (in press) found that mothers produced more canonical variants of word-final consonants in ID speech to female infants than male infants. Moreover, Foulkes, Docherty, and Watt (2005) found that mothers produced more canonical variants of /t/ to female children than male children in both word-medial and word-final positions. With respect to age of the infants, Bernstein Ratner (1984) and Sundberg and Lacerda (1999) have suggested that mothers' hyperarticulation is tied to the linguistic development of the infant. According to a proposal by Sundberg and Lacerda (1999), for example, mothers' speech to infants may be hypoarticulated before age one, hyperarticulated around age one when children first show signs of robust linguistic development (Oviatt, 1980), and then again hypoarticulated as children's speech develops further. In order to evaluate this proposal, it is necessary to examine ID speech across a large developmental window in order to get a representative account of clarity in ID speech.

To address these issues, we examined VOT production in ID speech to infants at ages 0;3, 0;9, 1;1, and 1;8 in a large, cross-sectional study with 48 mother-infant dyads. Examining this age range, which spans critical points in infants' language development (Oviatt, 1980), in a large-scale study format allowed us to get a clearer picture of possible hyperarticulation in ID speech and to test the proposal of Sundberg and Lacerda (1999). We hypothesized that mothers' speech to infants should be clearest around 1;0, in line with the proposal that mothers' speech clarity is dependent on the infants' developmental trajectory. Given the previous findings showing that ID speech is clearer to female infants than male infants, we also expected to see this pattern in our data.

## METHODS

### Participants

Participants were 48 mother-infant dyads recruited through the DeVault Otologic Research Laboratory at Indiana University School of Medicine. All were native speakers of American English. Each mother was recorded reading to her infant during a single visit when the infant age was near 0;3 ( $M = 0;3.2$ ,  $SD = 0;0.13$ ), 0;9 ( $M = 0;9.1$ ,  $SD = 0;0.13$ ), 1;1 ( $M = 1;0.25$ ,  $SD = 0;0.16$ ), or 1;8 ( $M = 1;8.13$ ,  $SD = 0;0.27$ ). Of the 48 infants in the study, 29 were

male and 19 were female. All mothers were reimbursed \$10 per visit. This research and the recruitment of human subjects were approved by the Indiana University Institutional Review Board.

## Procedure

### *Materials and Apparatus*

In the ID speech condition, each mother was asked to read a storybook to her child and to sit with her child in the sound booth and interact with her child as she normally would at home. The text of the storybook is shown in the Appendix; the text was constructed to consist of a range of consonants and vowels of interest in controlled phonetic contexts. The storybook was illustrated with age-appropriate color pictures. In the AD speech condition, each mother was asked to read the same storybook aloud while she was alone in the sound booth as she would to an adult. AD session lengths ranged from 1.3 to 2.2 minutes ( $M = 1.7$ ;  $SD = 0.2$ ), and ID session lengths from 1.4 to 3.9 minutes ( $M = 2.6$ ;  $SD = 0.5$ ). The order of ID and AD recordings was roughly counterbalanced across mothers; twenty-one dyads participated in the AD session first, and the other 27 participated in the ID session first.

Recordings were made at the Indiana University DeVault Otologic Lab/Babytalk Lab in a double-walled, copper-shielded sound booth (Industrial Acoustics Company). Each mother's speech was recorded by a hypercardioid microphone (Audio-Technica ES933/H) that was powered by a phantom power source and suspended from the ceiling. This microphone was linked to an amplifier (DSC 240) and digital audio tape recorder (Sony DTC-690). The speech samples were recorded directly onto a Mac computer (Apple, Inc. OSX Version 10.4.10) via Hack TV (Version 1.11) software.

### *Data Analysis*

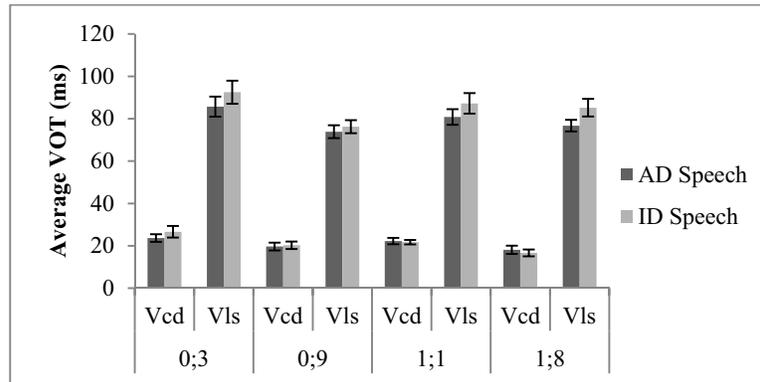
Tokens were individual productions of the voiceless stop consonants /p, t, k/ and the voiced stop consonants /b, d, g/ occurring in word-initial position in the storybook. Only tokens in which the stop consonant occurred in a primary stress syllable were analyzed, including extemporaneous productions of these words, such as might occur in phrases like "Look at the cute brown dog!"

Phonetic analysts were five students (undergraduate and graduate) who attended three training sessions and completed two practice assignments before being approved for involvement in the analysis. Each token was analyzed by a single analyst, who identified the release of the stop closure (burst) and the onset of voicing. Using Praat (Boersma and Weenink, 2006), analysis was based on waveform and spectral cues in conjunction with auditory percepts. The location of the burst was identified based primarily on the occurrence of a band of high-frequency energy in spectrogram that corresponded to the auditory percept of a stop release. When the location of the burst was unclear in the spectrogram, analysts referenced the waveform for evidence of aperiodic energy corresponding to the auditory percept of a stop release. The location of the voicing onset was identified based primarily on information from the waveform, and was marked at the beginning of the first full pitch period, i.e., a positive plus negative going that resembled the next period in the waveform and corresponded roughly with the appearance of a voicing bar and formants in the waveform. In instances of continuous voicing from the previous word (i.e., periodicity in the waveform continued from a preceding segment through the stop closure), voicing onset was marked on a positive-going zero-crossing at the exact location of, or immediately following, the burst. Tokens were excluded if the mothers' speech was unclear (whispered, glottalized, and/or accompanied by the baby), or if the spectrogram did not give definitive evidence of both a burst and onset of voicing, possibly due to approximation or soft speech. Moreover, prevoiced segments were excluded from the analysis because of their low frequency (<2% of all voiced tokens for both AD and ID speech).

Inter-rater reliability across analysts was assessed by measuring variability across all analysts' measurements of VOT for a single mother's AD and ID speech, for a total of 181 measurements per analyst. The inter-rater reliability analysis revealed very low standard deviations across tokens, indicating high reliability. The average standard deviation for voiced consonants was 0.8 ms, and the average standard deviation for voiceless consonants was 1.2 ms. VOT was calculated by subtracting the start time of the burst from the start time of voicing. (Note that there were no pre-voiced tokens included in our analysis.)

## RESULTS

A total of 8725 stop consonants were included in the analyses (4187 from AD speech, 4538 from ID speech). VOT is shown in Figure 1 as a function of Voicing, Infant Age, and Addressee. To ensure that mothers' speech was not affected by the order in which they read the book (AD-first vs. ID-first), an initial 5-way ANCOVA was run with VOT as the dependent variable and factors of Addressee (ADS, IDS), Voicing (Voiced, Voiceless), Place of Articulation (Labial, Alveolar, Velar), Infant Age (0;3, 0;9, 1;1, 1;8), and 2 Infant Gender (Male, Female) with the covariate of Order (ID-first, AD-first). Due to the finding that Order did not significantly affect VOT production, this variable is not discussed in the below results.

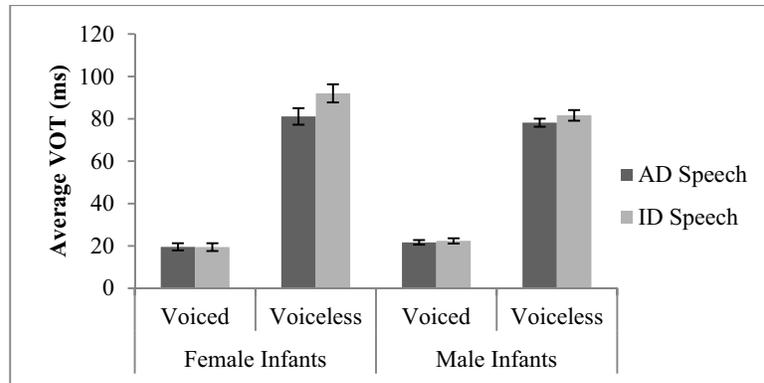


**FIGURE 1.** Average VOT (in ms) of voiced (“Vcd”) and voiceless (“Vls”) consonants in Adult-Directed (AD) and Infant-Directed (ID) speech as a function of Infant Age (years;months). Error bars are +/- 1 standard error.

A 5-way ANCOVA was performed on Addressee (ADS, IDS) X Voicing (Voiced, Voiceless) X Place of Articulation (Labial, Alveolar, Velar) X Infant Age (0;3, 0;9, 1;1, 1;8) X Infant Gender (Male, Female) with Prior Sessions (0-3) as a covariate and VOT as the dependent variable. Addressee, Voicing, and Place of Articulation (POA) were within-subjects variables, while Infant Age and Infant Gender were between-subject variables. The number of Prior Sessions was included as a covariate to ensure that any effect of familiarity was controlled for; this was because the speech for this study was taken from a large corpus of recorded materials and some mothers had come in for sessions prior to the interval included in this study.

As expected given the place and voicing contrasts that exist in English, Voicing ( $F(1,39) = 799.666, p < .001$ ) and POA ( $F(2,38) = 83.053, p < .001$ ) were significant as was the interaction between Voicing X POA ( $F(2,78) = 6.631, p = .002$ ); VOTs for voiceless consonants were significantly longer than their voiced counterparts (82 ms vs. 21 ms, respectively). Moreover, VOTs for velar consonants (68 ms) were longer than alveolar consonants (60 ms), which were in turn longer than labial consonants (40 ms), consistent with prior research (Lisker and Abramson, 1967).

Addressee was also significant ( $F(1,39) = 5.463, p = .025$ ), as was the interaction between Addressee X Voicing ( $F(1,39) = 12.064, p = .001$ ); VOTs were significantly longer in ID than AD speech (60 ms vs. 56 ms), specifically for voiceless consonants (85 ms vs. 79 ms). There was no significant difference in VOT length for voiced consonants in ID and AD speech (21 ms vs. 21 ms). Furthermore, while Gender did not reach significance ( $p = .116$ ), the interaction between Addressee X Voicing X Gender was significant ( $F(1,39) = 9.444, p = .004$ ), reflecting the Addressee X Voicing interaction being driven by the mothers of Female infants (see Figure 2.)



**FIGURE 2.** Average VOT (in ms) of voiced (“Vcd”) and voiceless (“Vls”) consonants in Adult-Directed (AD) and Infant-Directed (ID) speech as a function of Infant Gender. Error bars are +/- 1 standard error.

Infant Age was significant ( $F(3,39) = 3.057, p = .039$ ). However, no interactions between Infant Age and Addressee were found to be significant, suggesting that mothers do not systematically change their production of VOTs based on the infants’ ages during the age intervals in this study.

## DISCUSSION

This study investigated VOT production in a large, cross-sectional sample of 48 mothers-infant dyads in which infants were approximately 0;3, 0;9, 1;1, or 1;8. While a number of studies of VOT have been reported in the ID speech literature (e.g., Baran *et al.*, 1977; Malsheen, 1980; Sundberg and Lacerda, 1999), none have looked at a large sample of infants across a wide range of ages significant for linguistic development.

Across the four infant age groups, VOTs were significantly longer in ID speech than AD speech. Replicating previous findings (Malsheen, 1980), VOTs were significantly longer for voiceless consonants in ID speech than AD speech, while the VOTs for voiced consonants were not affected by the addressee. Moreover, this effect of voicing appears to be driven by the gender of the infant; while gender of the infant did not significantly affect the length of VOTs for ID speech overall, VOTs of voiceless consonants only reliably differ between ID and AD conditions in speech directed to female infants.

With respect to VOT production across development, there was no apparent change in mothers’ production of VOTs based on the infant’s age. This finding runs contrary to the proposal that hyperarticulation should peak around age one (Sundberg and Lacerda, 2001).

## CONCLUSION

In conclusion, the present study showed that mothers appear to articulate consonants more carefully when talking to infants than when talking to adults. Specifically, when producing word-initial stop consonants, mothers increase the perceptual difference between voiced and voiceless consonants by producing longer VOTs for voiceless consonants in ID speech than AD speech. In other words, the temporal difference between voiced and voiceless consonants is significantly longer in ID speech than in AD speech. While we found no evidence for change over time in clarity of ID speech, there was an effect of the infant’s gender on VOT; for the time period we examined, mothers appear to hyperarticulate their speech to female infants, but not male infants. Given that clarity in infant directed-speech has been linked to improved language development (Liu *et al.* 2003; Tsao *et al.*, 2004; Kuhl *et al.*, 2005), these findings have important implications for developmental research.

## ACKNOWLEDGMENTS

This work was supported by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

## REFERENCES

- Baran, J. A., Laufer, M. Z. and Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics* **5**, 339-50.
- Bergeson, T. R., Miller, R. J. and McCune, K. (2006). Mothers' speech to hearing-impaired infants and children with cochlear implants. *Infancy* **10**, 221-40. doi: 10.1207/s15327078in1003\_2
- Bergeson, T. R. and Trehub, S. E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science* **13**, 72-5. doi: 10.1111/1467-9280.00413
- Bernstein Ratner, N. (1984). Phonological rule usage in mother-child speech. *Journal of Phonetics* **12**, 245-54.
- Bernstein Ratner, N. and Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics* **12**, 285-9.
- Boersma, P., and Weenink, D. (2006). Praat [Computer software]. Amsterdam: University of Amsterdam.
- Burnham, D., Kitamura, C. and Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science* **296**, 1435. doi: 10.1126/science.1069587
- Dilley, L., Millett, L., McAuley, J.D., Bergeson, T. (in press). Phonetic variation in consonants in infant-directed and adult-directed speech: The case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language*.
- Englund, K. (2005). Voice onset time in infant directed speech over the first six months. *First Language* **25**, 220-34. doi: 10.1177/0142723705050286
- Fernald, A. and Simon, T. (1984). Expanded intonation contours in mother's speech to newborns. *Developmental Psychology* **20**, 104-13. doi: 10.1037/0012-1649.20.1.104
- Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek, U. Jürgens and M. Papoušek (eds.), *Nonverbal vocal communication: comparative and developmental approaches*. Cambridge: Cambridge University Press.
- Foulkes, P., Docherty, G. J., and Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81(1), 177-206.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science* **277**, 684-86. doi: 10.1126/science.277.5326.684
- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., and Pruitt, J. (2005). Early speech perception and later language development: implications for the "critical period". *Language Learning and Development*, **1**, 237-264.
- Liu, H.-M., Kuhl, P. K. and Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science* **6**, F1-F10. doi: 10.1111/1467-7687.00275
- Lisker, L., and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, **10**, 1-28.
- Malsheen, B. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. Yeni-Komshian, J. F. Kavanagh and C. A. Ferguson (eds.), *Child phonology*. New York: Academic Press.
- Oviatt, S. (1980). The emerging ability to comprehend language: an experimental approach. *Child Development* **51**, 97-106. doi: 10.1111/j.1467-8624.1980.tb02514.x
- Papoušek, M., Papoušek, H. and Bornstein, M. H. (1985). The naturalistic vocal environment of young infants: On the significance of homogeneity and variability in parental speech. In T. Field and N. Fox (eds.), *Social perception in infants*. Norwood, N.J.: Ablex.
- Shockey, L. and Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica* **37**, 267-74. doi: 10.1159/000259996
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, **4**, 1-22.
- Sundberg, U. and Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica* **56**, 186-99. doi: 10.1159/000028450
- Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. *Lund Working Papers in Linguistics*, **49**, 148-151.
- Synnestvedt, A. (2010). *Voice onset time in infant-directed speech at two ages*. (Doctoral dissertation, University of Maryland).
- Tsao, F. M., Liu, H. M., and Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, **75**, 1067-1084.

## APPENDIX

“Look What I Found” by Brittnie and Heather  
(Note: Analyzed segments are shown in boldface.)

The sweet pink kitten went for a walk and saw the cool green turtle. The cool green turtle found a little green key. Who did it belong to? The cool green turtle wanted the sweet pink kitten to help in finding who the key belonged to. As they were walking, the sweet pink kitten saw a small green ball. The sweet pink kitten and the cool green turtle were not sure who it belonged to. They picked up the small green ball and the little green key and kept

walking. Along the way, the cool green turtle found a pretty blue crystal. Once again, the sweet pink kitten and the cool green turtle wanted to know who the pretty blue crystal belonged to. They picked up the pretty blue crystal along with the little green key and the small green ball and kept walking. Then they saw the cute brown dog. He looked very sad! The cute brown dog said, “Oh no! I have lost my little green key, my small green ball and my pretty blue crystal. I dropped them and cannot find them anywhere!” The sweet pink kitten and the cool green turtle were very happy that they found who the little green key, small green ball and pretty blue crystal belonged to. The cute brown dog wanted his things returned. The sweet pink kitten and the cool green turtle were glad to return them, and this made the cute brown dog very happy.