The Development of Distinct Speaking Styles in Preschool Children

Melissa A. Redford
University of Oregon, Eugene, OR

Christina E. Gildersleeve-Neumann
Portland State University, Portland, OR

Purpose: To examine when and how socially conditioned distinct speaking styles emerge in typically developing preschool children's speech.
Method: Thirty preschool children, ages 3, 4, and 5 years old, produced target monosyllabic words with monophthongal vowels in different social–functional contexts designed to elicit clear and casual speaking styles. Thirty adult listeners were used to assess whether and at what age style differences were perceptible. Children's speech was acoustically analyzed to evaluate how style-dependent differences were produced.
Results: The ratings indicated that listeners could not discern style differences in 3-year-olds' speech but could hear distinct styles in 4-year-olds' and especially in 5-year-olds' speech. The acoustic measurements were consistent with these results: Style-dependent differences in 4- and 5-year-olds' words included shorter vowel durations and lower fundamental frequency in clear compared with casual speech words. Five-year-olds' clear speech words also had more final stop releases and initial sibilants with higher spectral energy than did their casual speech words. Formant frequency measures showed no style-dependent differences in vowel production at any age nor any differences in initial stop voice onset times.
Conclusion: Overall, the findings suggest that distinct styles develop slowly and that early style-dependent differences in children's speech are unlike those observed in adult clear and casual speech. Children may not develop adultlike styles until they have acquired expert articulatory control and the ability to highlight the internal structure of an articulatory plan for a listener.

KEY WORDS: language acquisition, speaking styles, speech acoustics, speech production

Speaking styles are socially conditioned linguistic modes characterized by differences in syntactic complexity, lexical choice, phonological form, and the phonetic realization of speech. Social distance, social context, and listener feedback are among the factors that trigger style shifting. For example, when speaking to a researcher, who is also a stranger, adults typically adjust the phonetic realization of their utterances, adopting a speaking style that is more fully articulated (i.e., less reduced) than their default speaking style (Labov, 1972). Speakers also unconsciously adopt this clear speaking style when a listener signals comprehension difficulty through his or her back-channel behavior or when listeners are perceived as having a high risk of comprehension difficulty (cf. Giles, Coupland, & Coupland, 1991). Because of the speakers' apparent focus on listeners, their clear speech has been characterized as "listener-oriented speech" in the phonetics literature (e.g., Lindblom, Brownlee, Davis, & Moon, 1992).

Consistent with its listener-oriented characterization, studies show that clear speech, whether inadvertently or deliberately produced, is more intelligible than casual speech for normal-hearing, hard-of-hearing, and non-native English-speaking listeners (Bond & Moore, 1994; Bradlow, Torretta, Pisoni, & 1996; Ferguson & Kewley-Port, 2002; Picheny, Durlach, & Braida, 1985). These studies have also documented the acoustic changes that are responsible for the clear speech intelligibility benefit. Somewhat less work has focused on understanding how the speaker implements such changes, perhaps because style shifting in adult language is so rapid and unconscious that the problem may seem fairly trivial. The question of implementation may be more interesting when one considers style shifting from the perspective of acquisition.

If you have talked with 2-year-olds, you have probably observed that they do not control speaking styles in the way that adults do. For example, if you indicate to a 2-year-old that you have not understood her, she will likely repeat exactly what she said before in exactly the same way. Although repetition is a rudimentary repair strategy that indicates sensitivity to the listener (Alexander, Wetherby, & Prizant, 1997; Brinton, Fujiki, Loeb, & Winkler, 1986; Ferrier, Dunham, & Dunham, 2000), the subjective experience is that 2-year-olds do not change their speech from repetition to repetition. The perceived similarity between repetitions suggests that young children have only one speaking style, which raises the questions of when and how children develop distinct speaking styles. In the present study, we sought to answer these questions by examining how 3-, 4-, and 5-year-old children produce the same target words under clear and casual speaking conditions.

Adult Clear and Casual Speech

The assumption from the adult literature is that speakers shift from a casual to a clear speaking style by manipulating some basic control parameters. For example, a style-dependent change in a global timing variable is suggested by the finding that clearer, more intelligible speech is slower than casual, less intelligible speech (Picheny et al., 1985; Picheny, Durlach, & Braida, 1986). An overall slowing in clear speech may lead to secondary changes in segmental articulation: Slowing the sequential execution of articulatory movements decreases the likelihood of segmental overlap (coarticulation) and increases the likelihood of attaining sequentially specified articulatory/acoustic targets (Gay, 1981; Lindblom, 1963; Moon & Lindblom, 1992; Munhall, Kawato, & Vatikiotis-Bateson, 2000). A temporally induced decrease in coarticulation may explain why the clear speech vowels are more dispersed in F1 × F2 space than casual speech vowels (Ferguson & Kewley-Port, 2002; Moon & Lindblom, 1992; Picheny et al., 1986; Smiljanic & Bradlow, 2005). It may also explain why no such vowel space expansion is observed when style is manipulated but speech rate is controlled (Krause & Braida, 2004).

Clear speech also typically has more pauses than casual speech (Picheny et al. 1985, 1986). Although pausing is consistent with a global slowing of speech, pauses are not likely to be randomly dispersed in clear speech given the increased intelligibility of this style. Rather, the pauses are likely prosodic in nature, serving to highlight linguistic boundaries (Cutler & Butterfield, 1990; Frazier, Carlson, & Clifton, 2006).

Prosodic highlighting may also be relevant for explaining the effect of pitch on speech intelligibility. Bradlow et al. (1996) found that greater fundamental frequency (f0) ranges tended to correlate with higher intelligibility scores, regardless of the speaker's gender. If we assume that the f0 range is linked to pitch accenting such that a low accent is associated with the f0 minimum and a high accent with the f0 maximum, then the correlation between f0 range and intelligibility suggests a clear speech expansion of the pitch accent space. Such an expansion could be achieved by adjusting a global pitch range parameter at the beginning of an utterance.

An expanded pitch accent space may function like an expanded vowel space—namely, to increase the perceptual distance between phonological categories in clear compared with casual speech (Diehl & Lindblom, 2002; Hay, Sato, Coren, Moran, & Diehl, 2006; Johnson, Flemming, & Wright, 1993; Lindblom, 1990). Some researchers have noted that other paradigmatic contrasts are also enhanced in clear compared with casual speech. For example, word-initial voiceless stops have longer voice onset times (VOTs) in clear than in casual speech (Krause & Braida, 2004; Picheny et al., 1986; Smiljanic & Bradlow, 2008), increasing the salience of the English voicing contrast. Word-final stops are more often released in clear speech than in casual speech (Krause & Braida, 2004), increasing the number of place-of-articulation cues available for these stops. And the mean spectral energy (center of gravity [CoG]) is much higher for palato-alveolar fricatives in clear than in casual speech (van Son & Pols, 1999), meaning that sibilants are more distinct from nonsibilants in clear compared with casual speech.

Like the expansion of the vowel and pitch accent space, the greater distinction between consonantal categories in clear compared with casual speech might be implemented by manipulating some basic control parameters. For example, van Son and Pols (1999) suggested that the style-dependent CoG differences in their study might have been due to greater subglottal pressure in clear compared with casual speech. Increased subglottal pressure may reflect increased articulatory effort consistent with Lindblom's (1990) hypothesis that clear speech represents *hyperspeech*—a mode in which speakers overcome articulatory inertias to maximize segmental target attainment. More audible releases of word-final stops in clear compared with casual speech could be explained similarly—that is, in terms of greater source strength in addition to more supraglottal articuatory effort. In contrast, the style-dependent VOT differences could be explained by changes in the global timing variable that was referenced to explain vowel space expansion in clear compared with casual speech.

In sum, adults may style shift by manipulating an articulatory timing and effort parameter and prosodic emphasis. Adult style shifting is characterized by suprasegmental and segmental acoustic changes, including a slower speech rate and expanded pitch range in clear compared with casual speech as well as a larger vowel space and greater contrast between consonantal categories.

Developing Distinct Speaking Styles

The assumption from the adult literature on style shifting is that clear speech is listener oriented. This suggests that children must acquire the relevant social and pragmatic skills before they develop distinct speaking styles—that is, children must be aware that they can adjust their speech to help a listener understand what they are trying to say. Studies on communicative repair strategies in young children as well as on context- or listener-induced language switching in bilingual children suggests that this awareness is in place by the end of the second year of life (Alexander et al., 1997; Brinton et al., 1986; Ferrier et al., 2000; Genesee, 2001; Genesee & Nicoladis, 1997). Children as young as 3 years old may, therefore, be able to style shift, but whether or not they do so will depend in part on whether they have acquired adultlike control strategies.

We noted previously that adult style shifting likely results from changes to global articulatory parameters such as timing and effort. We might reasonably assume that even young children are able to adjust global articulatory timing and effort, given that they can speak more or less quickly and more or less loudly. It does not follow, however, that children would automatically make such articulatory adjustments in response to communicative demands or that the adjustments would have the same acoustic results as in adult speech.

With respect to the latter point, global articulatory adjustments could have different effects in child and adult speech because children's representations may differ from those of the adult. For example, if abstract phonemic knowledge is acquired slowly from generalization across continuous word level representation, as has been repeatedly suggested (Beckman & Edwards, 2000; Lindblom, 1992; McCune & Vihman, 2001; Storkel & Morisette, 2002; Ziegler & Goswami, 2005), then children are likely to have different phonemic representations than adults. Children's smaller vocabularies may result in phonemic representations with simplified internal structure compared with adult representations, which are generalized over much larger vocabularies. Insofar as a phoneme's internal structure represents paradigmatic contrasts, these contrasts may be weakly represented or absent in children's simplified phonemic representations. If children's phonemic representations do not code contrasts as stongly as adults', then children are unlikely to "enhance" such contrasts when switching to clear speech. This could mean that children's style shifting will not result in the vowel space expansion and consonantal category distinctions that are observed in adult clear speech if these are indeed attributable to phonemic contrast enhancement, as some have argued (Diehl & Lindblom, 2002; Hay et al., 2006; Johnson et al., 1993; Lindblom, 1990).

At the suprasegmental level, global articulatory adjustments to timing and pitch in adult speech result in more or less emphatic prosodic highlighting with the acoustic consequences of increased pausing and increased distance between low and high accents in clear compared with casual speech. Again, these acoustic consequences follow not only from articulatory adjustments but also from the structure of the underlying representation. Even if children slow their speech and adjust their pitch range to accommodate a listener, the acoustic consequences of these changes will only be adultlike if the relevant underlying prosodic representations are adultlike. This could mean that some suprasegmental aspects of children's clear and casual speech will differ from those of the adult for an extended period of time, given that some aspects of prosody are not acquired until late childhood (Wells, Peppé, & Goulandris, 2004).

In sum, children as young as 3 years old may be inclined to accommodate a listener by style shifting, but the different speech styles that they produce are unlikely to resemble adult clear and casual speech because children's phonemic and prosodic representations likely differ from adults' representations.

The Present Study

In the present study, we investigated when and how distinct speaking styles develop in preschool children, keeping in mind that style shifting in child speech may not resemble style shifting in adult speech. Our method was to create the ecological conditions that would naturally elicit clear and causal speech in an adult. Specifically, we manipulated social distance, context, and feedback in order to elicit different speaking styles in preschool children. In the clear speech condition, children were "tested" on their picture naming ability by an experimenter who was a stranger to the child and who instructed the child to speak as clearly as possible. In the casual speech condition, the experimenter was absent, and children were recorded while playing blocks with their caregiver.

The reader may note that our method of speech style elicitation differs sharply from the straightforward method used in the adult phonetics literature, which typically involves instructing speakers to speak "clearly" or "normally" (Cutler & Butterfield, 1990; Moon & Lindblom, 1992; Picheny et al., 1985, 1986). Our more elaborate method for eliciting distinct styles is necessary because we cannot assume that children have the same metalinguistic abilities as adults. In particular, the straightforward method used in adult studies assumes sufficient metalinguistic knowledge to access distinct, socially defined speaking styles in the absence of appropriate social or contextual cues. If children have not yet developed distinct speaking styles, it is not reasonable to assume that they would have acquired the metalinguistic abilities necessary to access such styles on demand. In other words, we assume that the acquisition of a meta-ability follows the acquisition of the actual ability, and so we tried to provide the full range of appropriate social and contextual cues necessary to elicit different speech styles.

Once the socially defined speech samples were obtained, adult listeners provided blind goodness ratings on matched words from the clear or casual conditions, attending either to suprasegmental or segmental attributes. The goal was to determine the age at which style-dependent differences are first perceptible and whether these differences are more notable at the suprasegmental or segmental level. A variety of phonetic measures were then taken to assess in more detail the specific character of style-dependent production differences. The overall goal was to understand the development of distinct speaking styles.

Method

Participants

Thirty typically developing American English–speaking preschool children participated in the cross-sectional developmental study. Ten children were 3 years old (3;4 to 3;7 [years;months]), ten were 4 years old (4;4 to 4;7), and ten were 5 years old (5;3 to 5;8). All children were being raised in monolingual, English-speaking households. Parents also confirmed that all children had normal hearing and had exhibited normal development. Normal development was determined by asking parents about the age at which their children had reached a number of well-known speech and motor milestones (e.g., age of first canonical babble, age of first steps).

In addition to the preschool children, 30 undergraduate students from the University of Oregon participated in the study. These students had self-reported normal hearing and were native speakers of American English. They provided clearness and adultlikeness ratings on the children's productions in exchange for course credit.

Speech Sample

Tasks. The experimenter welcomed the child and his or her parent into a small, quiet observation room. The experimenter, child, and parent then sat around a child-sized table, and the study began with the experimenter administering consent and obtaining information on developmental milestones. The child then completed the clear and casual speech tasks as well as an imitation task (for a separate study). The experimenter would typically administer either the imitation or clear speech task first. She would then leave the room, during which the child and parent would engage in the casual speech task. Upon returning, the experimenter would administer the remaining test-type task. This order was not always followed, though, as we focused more on sustaining a child's interest in the study than on counterbalancing the order in which tasks were completed.

As noted in the introduction, the preschool children participated in two tasks that were designed to elicit clear and casual speech, respectively. In the clear speech condition, the picture naming task was presented to the child as a test. The experimenter proceeded to ask the child whether or not he or she could name each of the objects shown on 5″ × 7″ laminated cards (see *Stimuli* subsection that follows). Children were recorded as they named the pictures that were presented to them one at a time in a randomized order. Although presented as a test, the task was simple for all children, and none had trouble spontaneously naming the pictures. In order to obtain maximally clear speech, the experimenter instructed children to speak clearly. The experimenter also provided intermittent feedback to ensure clear speech. For example, if the child was speaking softly, the experimenter coached him or her to speak louder, and if the child became silly or too informal, the experimenter coached him or her to use their "big boy" or "big girl" voice. The recordings were made using a high-quality standing microphone (Shure BG 5.1), which was placed in the center of the table, and a Tascam DA-P1 digital tape recorder or a Marantz PMD670 flash memory recorder. A standing microphone was used to create a formal recording environment. Its presence was noted, and speaking for the microphone was made an explicit part of the task.

In the casual speech condition, the experimenter left the room ostensibly to allow the child and parent time to play together. Before she left, the experimenter provided the dyad with a set of wooden blocks. Each of the blocks had a stimulus item pictured on two of its four sides. Parents were instructed to play some kind of game—categorization, building, or story-telling—that would encourage their child to spontaneously name the pictures on the blocks. The spontaneous speech produced while playing was recorded using the same equipment as that used during the picture naming task. The standing microphone was used in preference to a wireless microphone to maximize the continuity between conditions; however, in this condition, the recorder and microphone were placed to the side, away from the play area. Although the microphone remained directed at the child, parents were encouraged to ignore its presence. In this way, children were implicitly encouraged to view the equipment as furniture rather than integral to the task of playing with blocks.

Stimuli. The stimuli were pictures of familiar objects obtained from Boardmaker software (Mayer & Johnson, Inc., Pittsburgh, PA). Although 63 pictures were selected to broadly sample children's speech sound repertoires, the current study focused on the subset of pictured objects with monosyllabic names that had monophthongal syllabic nuclei without "r" coloration. Monosyllables were chosen to control for effects of lexical stress and foot structure on vowel production. Monophthongal syllabic nuclei without "r" coloration were chosen to simplify the formant frequency measures that were taken (see the paragraphs that follow).

The composition of the casual speech word set was further determined by the spontaneous productions in the casual speech condition. Although all children named some subset of the pictures on the blocks, the size of the subset that met the study criteria varied from child to child. Altogether, the children produced 25 different words that met the criteria. These words are shown in Table 1.

The monosyllabic words were spoken with similar illocutionary intent in the clear and casual speech conditions. In the casual speech condition, children often labeled the blocks as they played. This behavior was in response to parents' usually implicit and sometimes explicit encouragement to do so. Accordingly, just like in the clear speech condition, many of the target words produced in the casual speech condition were produced in isolation with or without a determiner, and many others were produced in utterance-final position. This pattern did not vary substatially with age: 3-year-olds produced 59.4% of target words in isolation and 23.4% in final position; 4-year-olds produced 72.9% in isolation and 17.8% in final position; and 5-year-olds produced 53.3% in isolation and 24% in final position. A smaller number of target words were excised from utterance-initial and utterance-medial position. The proportion of words excised from these two contexts varied with the type of play task in which the children and parents engaged, but not much with age: 3-year-olds produced 10.9% of target words in initial position and 6.3% in medial position; 4-year-olds produced 6.2% in initial position and 3.1% in medial position; and 5-year-olds produced 8% in initial position and 14.7% in medial position. The most voluble children told stories in which the target words were embedded in utterances. Interestingly, these children often uttered fewer of the monosyllabic words that are the focus of the present study, preferring instead to build their stories around "caterpillars," "giraffes," "elephants," "butterflies," and other entities with multisyllabic names that were excluded from these analyses.

Table 1. Pictured stimuli elicited the following monosyllabic words with monophthongal vowels.

Vowel type	Elicited words					
H.F.	cheese	key	leaf	pig	sheep	tree
M.F.	bed					
L.F.	bath	cat				
M.C.	bus	cup	drum	duck	sun	truck
L.B.	ball	clock	dog	sock		
H.B.	book	broom	juice	shoe	spoon	wolf

Note. The words are arranged by vowel type. H.F. = high front; M.F. = mid front; L.F. = low front; M.C. = mid central; L.B. = low back; H.B. = high back.

The monosyllabic target words that the child uttered while playing were matched to the target words produced during the formal picture naming task—that is, speech style was a within-subjects factor. Children would occasionally produce a target word more than once in the casual speech condition. When this happened, only the first word was chosen for comparison with the clear speech word so that only words with new information statuses were included in the analyses.

Because children varied in their volubility in the casual speech condition and inevitably focused on different objects to name, the analyses were based on different sets of the monosyllabic target words for each child. The number of matched clear and casual speech word pairs produced also varied somewhat with age group. The ten 3-year-old children produced the fewest matched words, with a mean of 6.6 pairs per child, a range of between 3 and 12 pairs, and an *SD* of 2.95 pairs. The ten 4-year-old children produced the most matched words, with a mean of 12.9 pairs, a range of between 8 and 18 pairs, and an *SD* of 3.45 pairs. The ten 5-year-old children produced a mean of 7.7 matched words, with a range of between 4 and 14 word pairs and an *SD* of 3.3 pairs. Overall, the 30 preschool children produced a total of 272 matched pairs or 544 clear and casual speech words.

Although there was structural variability in the target words, the analyses of style were within subject, so clear and casual speech words were always matched for syllable structure as well as for vowel and consonant type. Our inferential statistics were based on the restricted maximum likelihood method of estimating variance and covariance, a procedure made necessary by the different number of matched word pairs per child.

Listener Ratings

The 544 clear and casual speech words were combined and presented in random order to three groups of

10 adult listeners, who rated word productions on nine-point Likert scales according to specific instructions. Listeners were linguistically and phonetically naïve, as their task was merely to use their own native English-speaking judgments to rate children's productions. This is analogous to asking undergraduate and graduate students, who have had no formal training in the theory or practice of education, to rate instructor "effectiveness" on course evaluations.

The goal of the rating task was to evaluate, in an omnibus fashion, the effects of age and speaking style on children's word productions. The specific instruction conditions were used to focus listeners' attention on different subsets of production attributes. In particular, one group of listeners was instructed to rate the clearness of whole word production, another group to rate the clearness of vowel production, and a third to rate the adultlikeness of the word production. Clearness ratings were obtained to evaluate suprasegmental and segmental attributes of the children's word productions. Adultlikeness ratings were obtained to evaluate cues that differentiate child and adult speech under the assumption that these might influence listeners' specific metalinguistic notions of clear versus casual speech production.

The rating task was completed one listener at a time in a sound-attenuated booth. A listener received a list of all the target words that they would hear in the predetermined randomized order in which they would hear them. Each word was listed with a nine-point number scale—a scale sensitive enough to capture the presumably small differences between age groups and speaking styles in young children's speech (see, e.g., Southwood & Flege [1999] for a discussion of the relative sensitivity of different Likert scales). In the adultlikeness condition, the nine-point scale was anchored at the top of each page with the words "least adultlike" and "most adultlike" appearing above numbers 1 and 9, respectively. Listeners were instructed to rate the adultlikeness of all the words they heard, using as much of the scale as possible. In the whole word and vowel conditions, the nine-point scale was anchored at the top of each page with the words "least accurate" and "most accurate." "Most accurate" was defined for the listeners as either (a) the clearly articulated adult version of the target word or (b) the vowel in a clearly articulated adult version of the target word. The particular definition corresponded to whether a listener was instructed to attend to the whole word or the vowel. "Least accurate" was not defined. Instead, listeners were encouraged to use as much of the scale as possible when making their judgments.

Listeners were quite consistent in their relative rating of individual words. The Cronbach's alphas were .89, .83, and .88 for listeners in the word, vowel, and adultlikeness instruction conditions, respectively. However, listeners were highly variable in their absolute rating of individual words. The Shrout-Fleiss coefficients were .23, .26, and .17 for those in the word, vowel, and adultlikeness conditions. The high Cronbach's alphas and low Shrout-Fleiss coefficients suggest that listeners assessed accuracy using different parts of the nine-point scale or different ranges within the scale.

The highly variable listener ratings coupled with the nine-point scale provided sufficient variation in the data to approximate a continuous response function. This approximation allowed us to examine the effects of random and fixed factors in a mixed univariate model that included all listener judgments (N = 16,227). The results from this analysis were qualitatively similar to those that used an average per item score. The difference between the two types of analyses was in the effect sizes. The effect size was diminished in an analysis based on averages because so much data are lost in averaging across highly variable judgments.

Acoustic Measures

The preschool children's 544 words were also acoustically analyzed. The measures focused on suprasegmental and segmental attributes associated with vowel production and some additional segmental attributes associated with consonant production. The vowel segmentation criteria and measures are described first.

The recorded speech was displayed as oscillograms and spectrograms simultaneously in Praat (Boersma & Weenink, 2007). Target word onsets and offsets were typically identified by pauses, but words were sometimes extracted from phrases in the casual speaking condition (see previous *Listener Ratings* section). Vowel onsets and offsets within the word were identified from amplitude and periodicity changes in the waveform. Auditory judgments were used to confirm the visual segmentations, and absolute word and vowel durations were taken. Relative vowel duration was then calculated as the ratio of vowel duration to word duration. To ensure that vowel segmentation was reliable, a second rater independently re-measured all vowel durations, and a single-measure intraclass correlation coefficient (ICC) was calculated, showing good interrater reliability (ICC = .85).

In addition to the duration measures, pitch measures were taken to evaluate style-dependent suprasegmental changes. These measures included the mean f0 value for the total vowel duration as well as the f0 maximum and minimum over this period. The *SD* around the mean was also recorded for each vowel. All measures were obtained automatically using the voice report function in Praat.

The segmental measures focused on vowel and sibilant target attainment and on stop release and VOT. The vowel measures are described first. Several formant measures were taken for the vowel—namely, F1 midpoint, F2 onset, and F2 midpoint. The formant measures were taken by hand from visual inspection of the spectrogram, an inspection that was supplemented by the formant tracking (linear predictive coding [LPC]) function in Praat. The formant tracking parameters were set so that the algorithm searched for a maximum of four formants under 5500 Hz. The built-in peak picking algorithm was not used for measuring formants because the tracks were often discontinuous, which means that at vowel midpoint, the algorithm would return noise or values associated with higher or lower formants. To ensure reliability of the hand measurements, a second rater independently re-measured all 544 vowels. The interrater reliability was high: Single-measure ICCs for F1 midpoint, F2 onset, and F2 midpoint were .89, .83, and .89, respectively. It is also worth noting that the F1 and F2 midpoint measurements obtained here were consistent with previously published data on American English–speaking children's vowels (e.g., Lee, Potamianos, & Narayanan, 1999).

The formant measures yielded four dependent variables for analysis of style-dependent changes to vowel articulation: F1 midpoint, F2 midpoint, the difference between F1 and F2 (F2 – F1), and the extent of the F2 onset transition in Hz. This latter variable was calculated by subtracting the F2 onset value from the F2 midpoint value. The resulting difference was transformed so that analyses were conducted only on the unsigned values.

The formant measures were used to assess segmental target attainment as a function of speaking condition. In order to evaluate whether consonant production varied by speaking condition, we measured VOT for word-initial stop consonants and the mean spectral CoG for word-initial sibilants. The VOT measures were taken for voiced and voiceless singleton stops from the burst to the onset of voicing. Negative VOTs were combined with 0 and positive VOTs in the analyses. The CoG measures were taken for the duration of the fricative and were obtained using the automatic function in Praat. Because the consonantal measures were added late in the study, intrarater rather than interrater reliability estimates were obtained for the VOT measures. Specifically, 25% of the word-initial stops were randomly selected and remeasured by the same rater, who was blinded to the initial measurements. Cronbach's alpha was calculated for the two measurement sets, and the calculation indicated excellent intrarater reliability (α = .99). CoG measurements were not re-measured, as these had been obtained automatically and so were less susceptible to human error.

Finally, the number of final consonant deletions, consonant cluster simplifications, and final stop releases were also recorded. Judgments on these phonological variables were made by listening to the productions and by inspecting the acoustic waveforms for evidence of deletion, epenthesis, and stop releases. Whereas the results showed that 59.3% of all final stops (N = 268) were released, only 10.1% of all final consonants (N = 446) were deleted (14.8% in 3-year-olds' speech, 8.3% in 4-year-olds' speech, and 9.0% in 5-year-olds' speech) and only 5.6% of all clusters (6 out of 108) were simplified through consonant deletion (by 4-year-olds) or epenthesis (by a single 3-year-old). Because of the low frequency of final deletion and cluster simplification, the results reported in the next section do not include analyses of these variables as a function of age and style.

Results

The adult listener rating results indicated that distinct speaking styles develop slowly over time. Style-dependent differences in rating scores increased with age. The rating results and acoustic analyses also suggested that style differences first manifest as differences in timing and pitch. By the time a child is 5 years of age, however, style differences also manifest as the differential attainment of some segmental targets. In the section that follows, the detailed adult listener rating results are presented first.

Listener Ratings

Listener ratings provided an omnibus measure of how children's speech differed as a function of age and speaking condition. Different instructions were used to direct listeners' attention to suprasegmental and segmental attributes. In particular, listeners were asked to focus either on the adultlikeness of children's productions or on the clearness of their whole words and vowels. Vowel-focused listeners were expected to focus largely on cues to vowel identity. Word-focused listeners were expected to focus on syllable structure attainment and on cues to vowel and consonant identity. Finally, listeners who rated adultlikeness were expected to attend to additional suprasegmental attributes, such as speech rate and pitch patterns, that might also influence adult judgments of clearness.

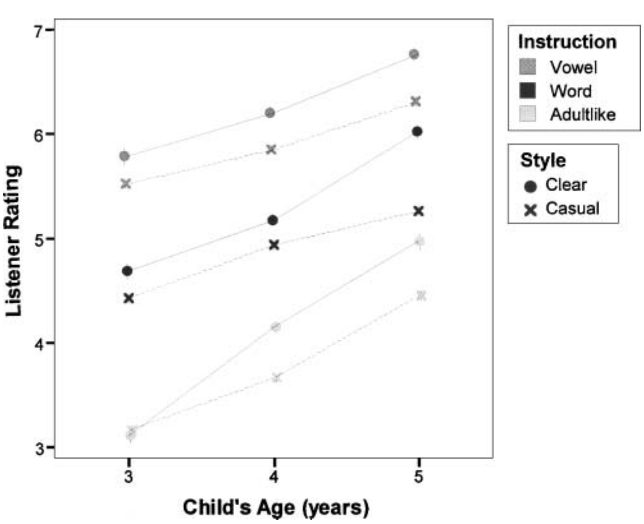
Listeners' rating scores were analyzed in a mixed model using a restricted maximum likelihood procedure for estimating variance and covariance. Children were treated as a random factor and were nested within age group (3, 4, 5), which was a between-subjects variable. Listeners were nested within instruction condition (vowel, whole word, adultlikeness) and then within age group. Speaking condition (clear or casual) and consonant number (a proxy for syllable structure) were the other within-subjects factors.

The aim of the first analysis was to determine whether or not children's speech differed in the clear or casual speech condition. The rating results indicated that it did: Listeners gave clear speech words higher average ratings than casual speech words, $F(1, 27) = 13.71$, $p = .001$, although this effect interacted with instruction condition and age, Style × Age × Instruction, $F(4, 54) = 4.74$, $p = .002$. The significant three-way Instruction Condition × Age × Style interaction is explored further in the paragraphs that follow.

Figure 1 shows that the type of instruction that a listener received influenced how that listener rated words produced in the clear and casual conditions, $F(2, 27) = 15.38$, $p < .001$. In particular, vowel-focused listeners rated children's productions more highly than did word-focused listeners, and adultlikeness ratings were the lowest of all. These differences between instruction condition might be explained to result from listeners attending to different numbers of production attributes, depending on their instruction condition, with vowel-focused listeners attending to the fewest attributes and adultlikeness listeners attending to the most.

Although there may have been some overlap in what differently-oriented listeners attended to, the significant Instruction × Age × Style interaction shown in Figure 1 suggests that groups of listeners also attended to differently weighted combinations of attributes. Three specific patterns illustrate this point. First, mean comparisons indicate that listeners did not differentiate between 3-year-olds' clear and casual words, although the comparisons show a trend toward style-dependent differentiation of 3-year-olds' words among vowel- and word-focused listeners: vowel, $t(54) = -1.55$, $p = .126$; word, $t(54) = -1.59$, $p = .118$. In contrast, mean comparisons showed that only the adultlikeness group differentiated between 4-year-olds' clear and casual words: vowel, $t(54) = -1.83$, $p = .072$; word, $t(54) = -1.31$, $p = .194$; adultlike, $t(54) = -2.47$, $p = .017$.[1] Third, listeners who focused on word-level attributes best distinguished between 5-year-olds' clear and casual words, even though all groups differentiated between the clear and casual words produced by this age group: vowel, $t(54) = -2.51$, $p = .015$; word, $t(54) = -3.92$, $p < .001$; adultlike, $t(54) = -2.85$, $p = .006$. Moreover, the style-dependent difference (d) in whole word ratings of 5-year-olds' speech was the largest effect of style overall: $d = .414$ (word) versus $d = .301$ and $.265$ (adultlike and vowel, respectively) and $d < .265$ for all mean comparisons on clear and casual words for all listener groups attending to 3- and 4-year-olds' productions.

Figure 1. The Instruction Condition × Speaking Style × Age interaction is shown for listener ratings. Ratings were on a 9-point scale (9 = *most accurate [adult clear speech] or most adultlike*, 1 = *least accurate or least adultlike*).



To better understand the large effect of style on the whole word ratings of 5-year-olds' speech, word ratings were reanalyzed as a function of age, style, and consonant number (cons_no). Consonant number substituted for syllable structure (see explanation that follows), a suprasegmental structural factor that may have affected children's word productions. The analysis confirmed that whole word ratings interacted systematically with all three of these factors, $F(4, 54) = 4.33$, $p = .004$. All two-way and simple effects were also significant: age, $F(1, 27) = 55.72$, $p < .001$; style, $F(1, 27) = 24.38$, $p < .001$; cons_no, $F(2, 27) = 11.24$, $p < .001$; Age × Cons_no, $F(4, 54) = 2.78$, $p = .036$; Style × Cons_no, $F(2, 54) = 6.22$, $p = .004$. The significant three-way interaction is shown in Figure 2.

As noted previously, consonant number corresponded with syllable structure (see Table 1). Words with one consonant had a CV structure. Words with two consonants had a CVC structure. Words with three consonants had a CCVC or CVCC structure. Figure 2 shows that listeners rated 3- and 5-year-olds' productions differently depending on syllable structure. In particular, the figure shows that 5-year-olds' productions of CCVC or CVCC words in the clear speech condition were rated as much more similar to adult clear speech versions of these words than were the 5-year-olds' productions of CCVC and CVCC words in the casual speech condition. The effect of style was somewhat smaller for 5-year-olds' productions of CVC and CV words. The Style × Consonant Number interaction was less predictable for 3-year-olds' speech. For example, 3-year-olds' casual

[1]It might be noted that $p = .017$ does not represent a significant difference if alpha is adjusted downwards with the number of comparisons being made. When alpha is adjusted in this way, significance is attained only when $p < .006$. However, there is disagreement in the literature as to whether this kind of adjustment is necessary. We are reporting results based on an unadjusted alpha on the assumption that a Bonferroni correction represents an overly conservative approach to the analyses (Perneger, 1998).

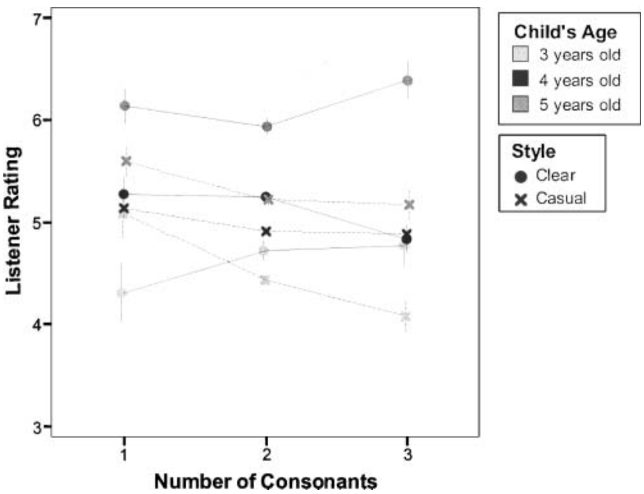


Figure 2. The Speaking Style × Age × Consonant Number interaction is shown for listener ratings in the whole word condition. Ratings were on a 9-point scale (9 = *most accurate [adult clear speech]*, 1 = *least accurate or least adultlike*).

speech productions of CV words were rated as more similar to adult clear speech versions of these words than were their clear speech productions.

Acoustic Measures: Suprasegmentals

Based on the rating results, our expectation was that style-dependent differences would be evident at the suprasegmental level in 4- and 5-year-olds' speech. This expectation was upheld in our analysis of duration and f0, both of which were found to vary systematically with the speaking style condition. The detailed results on duration are presented first.

Vowel duration is typically longer in adults' clear speech than in adults' casual speech (Moon & Lindblom, 1992; Picheny et al., 1986). Analyses of children's productions also indicated an effect of style on absolute vowel duration, but the direction of the effect was opposite from what is reported in the adult literature: Absolute and relative vowel durations were shorter in clear speech words than in casual speech words. These results are shown in Figures 3 and 4 and are discussed further in the text that follows.

Figure 3 shows a main effect of style on absolute duration, $F(2, 27) = 5.33$, $p = .029$: Casual speech vowels were longer than clear speech vowels. Vowel height also had a significant effect on absolute vowel duration, $F(2, 466) = 4.79$, $p = .009$. Low vowels were longer than mid and high vowels. In addition to these two main effects, there was a trend for vowel duration to vary predictably by age, $F(2, 27) = 3.04$, $p = .065$. In particular, 3-year-olds produced longer vowels than did 4- and 5-year-olds, but 3-year-olds also showed the most variability in vowel duration. The

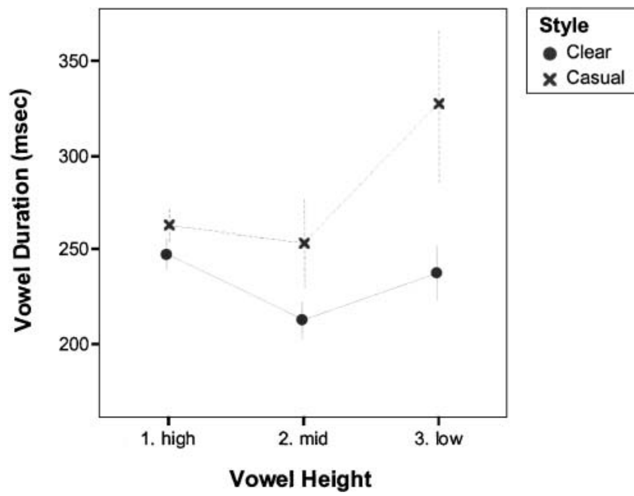


Figure 3. The Speaking Style × Vowel Height interaction is shown for absolute vowel duration.

variability of 3-year-olds' productions probably accounts for why the effect of age did not reach statistical significance. Neither style nor vowel height interacted with any other factor in the analysis.

Figure 4 shows the results on style-dependent differences in relative duration as a function of the child's age. Overall, casual speech vowels were longer relative to the word than clear speech vowels: style, $F(1, 493) = 6.50$, $p = .011$. In addition to the effect of style, the analysis showed an effect of age on the ratio of vowel duration to word duration: age, $F(2, 27) = 6.05$, $p = .007$. Three- and 5-year-old children produced longer vowels relative to the word than did 4-year-old children. Because the

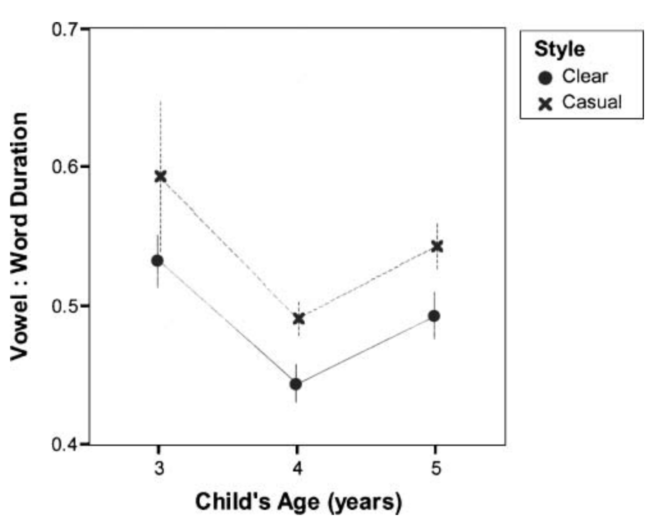


Figure 4. The Speaking Style × Age interaction is shown for relative vowel duration. Relative vowel duration equaled the absolute vowel duration divided by the absolute word duration.

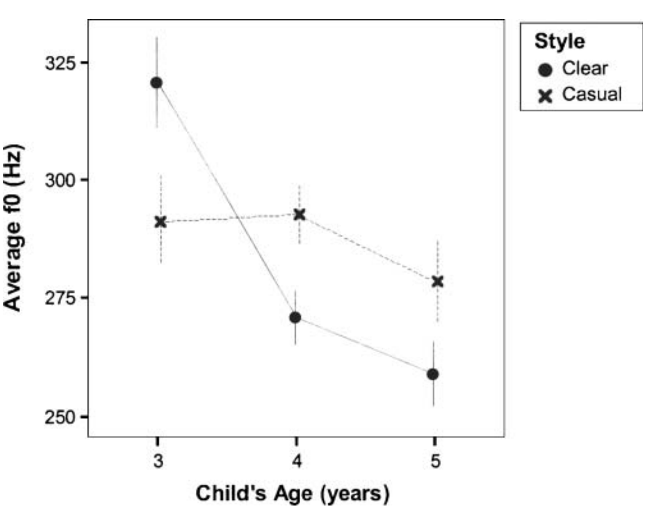
analysis of absolute vowel duration did not indicate a similar effect of age, the age-dependent differences in relative duration were probably due to differences in consonantal duration (i.e., word duration minus vowel duration). Thus, Figure 4 may be interpreted to show that overall consonantal duration was longer in 4-year-olds' speech than in either 3- or 5-year-olds' speech. This particular age-dependent difference did not interact with the effect of style, Style × Age, $F(2, 493) = 0.08$, $p = .927$; however, the figure indicates that, once again, 3-year-olds' productions were the most variable, and so the relative vowel durations overlapped in their clear and casual productions.

Pitch range also varies in adult clear and casual speech (Bradlow et al., 1996), but the f0 measurements indicated that this was not the case for children's clear and casual productions. In particular, analyses of the per-item f0 range (maximum minus minimum) and per-item *SD* around the f0 mean showed that neither measure was significantly affected by style.

Although dispersion around the mean was the same regardless of speaking condition and children's age, mean f0 varied systematically with these two factors, Style × Age, $F(2, 27) = 4.06$, $p = .029$, as shown in Figure 5. The simple effect of age was also significant, $F(2, 27) = 4.74$, $p = .017$, but neither style nor age interacted with any other factors.

Figure 5 shows that the significant Style × Age interaction on mean f0 was probably due to a developmental change that occurred between ages 3 and 4. Three-year-olds produced vowels with higher f0s in the clear speech condition than in the casual speech condition, but the reverse was true for 4- and 5-year-olds. Figure 5 also suggests, though, that 4-year-olds' clear and casual speech words were better distinguished by f0 than were 5-year-olds' words.

Figure 5. The Speaking Style × Age interaction is shown for average f0.



Acoustic Measures: Segmentals

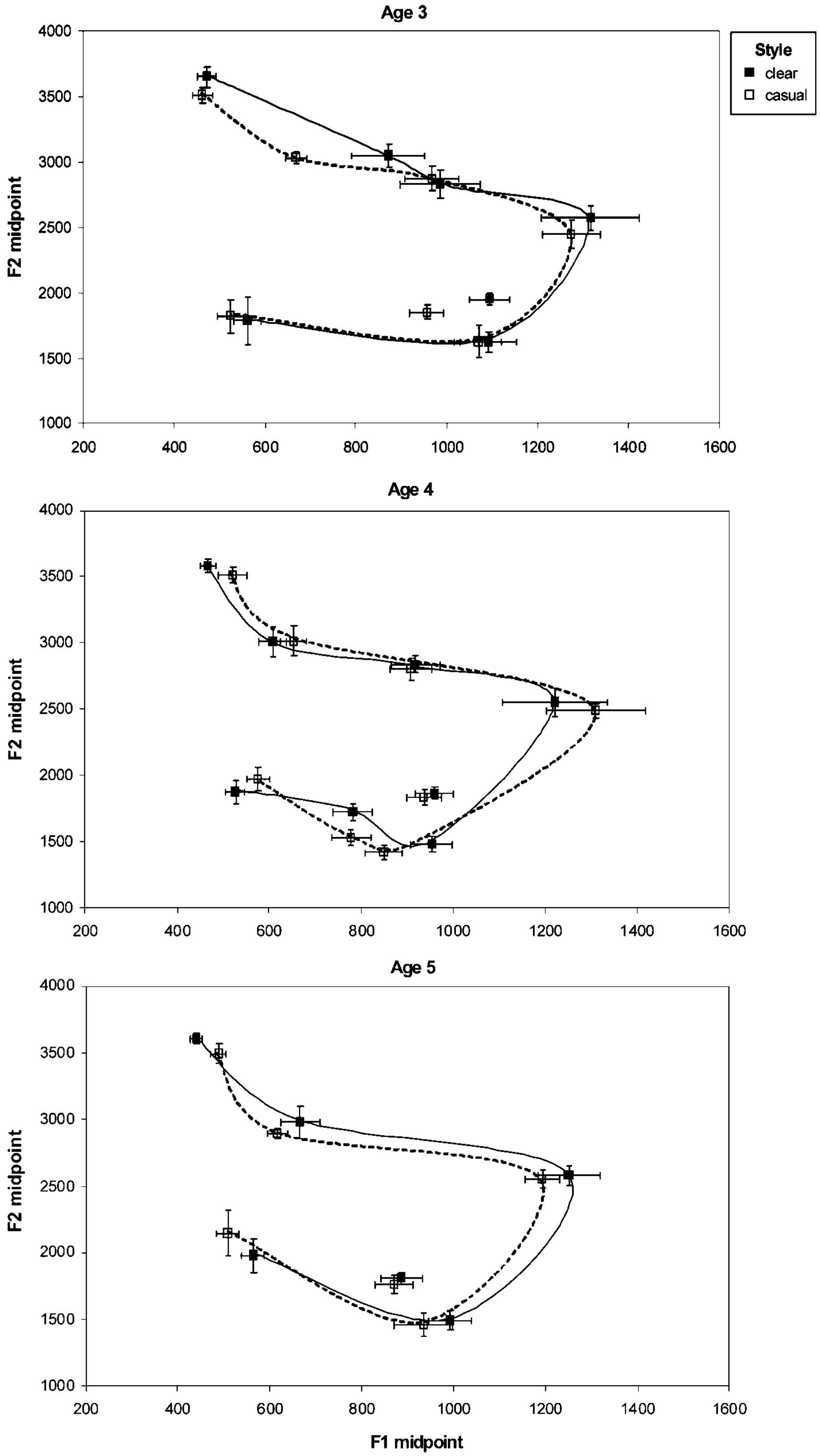
The second set of acoustic analyses focused on the segmental measures corresponding to target attainment for vowels and consonants. Recall that mean comparisons of the rating results indicated that word- and vowel-focused listeners reliably distinguished between the clear and casual productions of 5-year-olds but not between those of 3- and 4-year-olds. Accordingly, we expected that 5-year-olds would show reliable differences in segmental target attainment as a function of speaking condition. This expectation was not met for vowel production but was met for some aspects of consonantal production.

Vowels. Vowels in adult clear speech are more distributed in an F1 × F2 vowel space and exhibit more F2 displacement compared with vowels in adult casual or lower intelligibility speech (Bradlow et al., 1996; Ferguson & Kewley-Port, 2002; Krause & Braida, 2004; Moon & Lindblom, 1992; Picheny et al., 1986; Smiljanic & Bradlow, 2005). An analysis of F1 and F2 midpoints and F2 displacement in children's speech did not show these same effects.

Figure 6 shows the clear and casual vowels in an F1 × F2 vowel space for children, aged 3, 4, and 5 years, respectively. The estimated maximum likelihood procedure was used to assess the random effect of child, the between-subjects effect of age, and the within-subjects effects of style and vowel type (or height for F1 and backness for F2) on the formant measures. The results confirm what is evident from the figure—namely, that vowels were produced with similar F1 and F2 values regardless of speaking condition. Specifically, our measure of acoustic distance (F1 – F2) varied with vowel type, $F(5, 91) = 423.25$, $p < .001$, but not with style or any interaction between style and the other factors. F2 only varied systematically with vowel backness, $F(2, 47) = 586.29$, $p < .001$, and F2 transition only varied systematically with age and vowel backness, $F(4, 47) = 2.76$, $p = .039$. F1 varied systematically with vowel height, $F(50, 278.41) = 278.41)$, $p < .001$, and with style, $F(1, 27) = 4.34$, $p = .047$, but not with the interaction of these factors. The effect of style on F1 was, therefore, not related to changes in the distribution of vowels with respect to F1 but rather to an overall raising of F1 in the clear compared with casual condition.

Consonants. In adult clear speech, voiceless stop consonants are produced with longer VOTs, final stops are more likely to be released, and all consonants—especially sibilants—are produced with more energy in the higher frequencies (Krause & Braida, 2004; Picheny et al., 1986; van Son & Pols, 1999). Analyses of word-initial stop

Figure 6. Average F1 and F2 values are plotted for the 3-, 4-, and 5-year-olds' vowels produced in the clear and casual speaking condition. Error bars show the standard error of the mean. The solid and dashed lines delineate the perimeter of the clear and casual vowel space, respectively.



VOTs, word-final stop releases, and the CoG of strident consonants showed some of these same clear speech effects, especially in 5-year-olds' productions.

Voiced and voiceless stops were categorically distinguished by VOT in the clear and causal speech of all preschool children. The mean VOT for voiced stops was –8 ms (SD = 52), +4 ms (SD = 43), and +9 milliseconds (SD = 38) for 3-, 4-, and 5-year-olds, respectively. The mean VOT for voiceless stops was +90 ms (SD = 61), +103 ms (SD = 56), and +102 ms (SD = 52) for 3-, 4-, and 5-year-olds, respectively. Due to the categorical differences in the realization of voiced and voiceless stops, we only investigated the effects of age and style on VOTs within the voiced and voiceless stop category. These analyses indicated no significant effects of either age or style on VOT for voiced or voiceless stops: age, voiced, $F(2, 50) = 1.63$, $p = .21$; age, voiceless, $F(2, 50) = 0.48$, $p = .622$; style, voiced, $F(1, 310) = 0.04$, $p = .840$; style, voiceless, $F(1, 310) = 0.51$, $p = .475$. The interactions between the two factors were also not significant for either stop category.

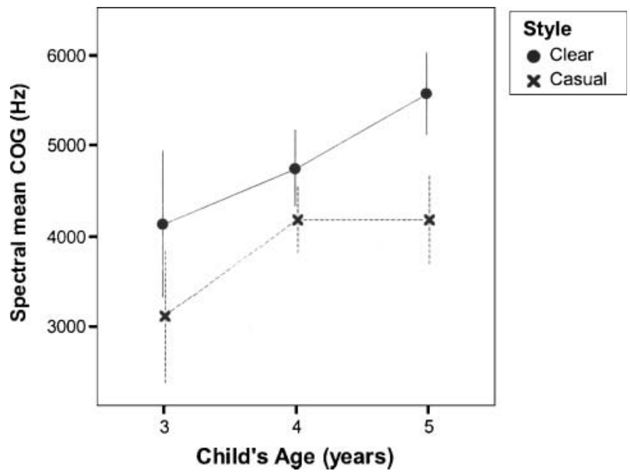
Although children's production of initial voiced and voiceless stops did not differ in the clear and casual speaking conditions, their production of final stops did, as measured by the presence or absence of a stop release. A hierarchical, random effects, logistic regression analysis showed that the number of final stop releases varied systematically with style, $F(1, 176) = 4.63$, $p = .033$. This effect of style is evident from the percent statistics reported in Table 2. These statistics clearly indicate that a larger percentage of final voiced stops were released in the clear compared with the casual speech condition across all age groups.

In addition to a style-dependent difference in the number of final stop releases, an analysis of word-initial sibilant CoG showed that the mean spectral center of gravity differed for clear and casual words, $F(1, 39) = 6.43$, p, = .015. Although this effect did not interact with any other factor, it is clear from the results shown in Figure 7 that there were differences across age groups. In particular, the figure suggests that sibilant CoG only varied systematically with style in the speech of 5-year-olds. This suggestion was supported by post hoc comparisons, which showed that sibilant CoG was not significantly different in the clear and casual words produced by 3- and 4-year-olds [3-year-olds, $t(39) = 1.23$, $p = .23$; 4-year-olds, $t(39) = 1.35$, $p = .18$] but that it was different in those words produced by 5-year-olds, albeit at a level that just misses statistical significance, $t(39) = 1.98$, $p = .054$.

Table 2. The percentage of word-final stops that were released is shown for the two speaking conditions, the three age groups, and the different stop types.

Age	Voiced stops (%)		Voiceless stops (%)	
	Clear	Casual	Clear	Casual
3 years (N = 20)	50.00	30.00	56.00	52.00
4 years (N = 18)	83.33	66.66	71.43	68.57
5 years (N = 12)	66.67	41.67	70.83	45.83

Figure 7. The Speaking Style × Age interaction is shown for the mean spectral energy (CoG) of word-initial sibilants.



Discussion

Overall, the results suggest that children acquire distinct speaking styles over several years in early childhood. The rating results indicate that listeners were unable to differentiate 3-year-olds' clear and casual word productions but were better able to differentiate between those produced by 4-year-olds and were especially able to distinguish between the clear and casual words produced by 5-year-olds. The rating results also suggest that, unlike 4-year-olds, 5-year-olds achieved the consonantal targets associated with different syllable structures better in clear than in casual speech.

The acoustic results were consistent with the rating results. These showed that 4- and 5-year-olds systematically produced words differently in the clear and casual speech condition. For 4-year-olds, the systematic differences were at the suprasegmental level and included differences in vowel duration and average f0. Like 4-year-olds, 5-year-olds differentiated clear and casual productions at the suprasegmental level, but they also showed emerging differences in their style-dependent productions of vowels and statistically significant differences in their style-dependent production of consonants.

Together, the rating and acoustic results provide some insight into how distinct speaking styles emerge. The results are consistent with the possibility that preschool children are able to adapt their speech to different social contexts by adjusting global articulatory parameters such as timing and pitch, even if the changes they make are different from the changes that adults would make under similar social circumstances. For example, the results on relative vowel duration in 4- and 5-year-olds' speech suggest style-dependent articulatory changes that are akin to rate changes in adult speech. Specifically, the longer relative vowel durations in 4- and 5-year-olds' casual speech words indicates vowel lengthening in the absence of significant consonant lengthening. This pattern parallels the nonlinear changes in segmental duration as a function of speech rate—vowel durations change more with speech rate than consonant durations (Gay, 1981)—and implicates a control strategy that relies on changes to a global timing parameter (see, e.g., Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985).

Apart from showing that children make global articulatory adjustments in response to different social contexts, the duration and f0 results indicate that such adjustments are qualitatively different than those that adults make under similar social conditions. Children produced faster speech in the clear compared with casual condition and kept the same pitch range settings in the two conditions but modified their register. These qualitative differences between child and adult speech in clear and casual speaking conditions could reflect differences in phonemic and prosodic representations, as suggested in the introduction. But the differences are probably also reflective of basic differences between children and adults—namely, differences in motor control, in default communicative postures, and in the dynamics of their social interactions.

Consider, for example, the notion of casual speech in the context of a developing motor system. Lindblom (1990) has characterized adult casual speech as the opposite of clear speech. Specifically, he has suggested that casual speech is *hypo-speech*—a mode in which the speaker defaults to a least effort strategy, resulting in more coarticulation between segments and so to the diminished distance between paradigmatic contrasts in acoustic space. The current results suggest that children's casual speech is not well characterized as hypo-speech. Consider the F1 × F2 distribution of vowels in Figure 6. This figure shows that children as young as 3 years old have a well-structured vowel space and that vowels are equally well distributed even when the social context would allow for articulatory reduction (or vowel undershoot). Children's consistently good production of vowels regardless of social context probably has less to do with deliberate attempts to achieve articulatory targets for the benefit of listeners than with an inability to take shortcuts when engaging in articulatory action. This idea is motivated by the finding that articulatory skills continue to develop throughout childhood (see Walsh & Smith, 2002), which suggests that children are not expert speakers. Insofar as novices use extensive and inefficient movements to obtain the same overall patterns that experts realize using efficient and reduced movements (e.g., Green, Moore, Higashikawa, & Steeve 2000; Green, Moore, & Reilly, 2002), it is likely that young children are simply incapable of the articulatory reduction that characterizes adult causal (hypo) speech. This possibility is consistent with kinematic data showing that young children make relatively larger amplitude speech movements than do adults (Riely & Smith, 2003).

Children's distinct speaking styles may also differ from adults' speaking styles because their default communicative posture may be different from an adult's in a casual speaking situation. A different default posture may provide the best account for the finding that absolute vowel duration was greater in children's casual speech than in their clear speech, which is the opposite of what has been reported for adult clear and casual speech. In the casual speech condition, children played with blocks and with sounds. Figure 3 shows that low vowels, in particular, were drawn out (e.g. "*A baaaaaalllll*"). Children inhibited this playful behavior in the clear speech condition, which explains why vowels were shorter. It may also explain why listeners rated clear speech productions as more adultlike than casual speech productions.

In fact, the adultlikeness ratings suggest that children were more serious overall in the clear speech condition when the task demands were to actively engage with an adult stranger. This may not be surprising, given that the stranger also repeatedly encouraged the child to be as clear and intelligible as possible. From a child's perspective, the social distance and corrective feedback (clearer, louder, less silly) may have been interpreted to mean that the child should speak like an adult. Such an interpretation may have been enforced here, just as it is in natural settings, by encouraging children to speak up and behave.

The idea that a child may understand clear speech to mean adult speech is also consistent with the f0 results, which contrasted with findings on adult clear and casual speech. Adult clear speech has an expanded f0 range, but children's f0 range did not change across speaking conditions. Instead, 4- and 5-year-olds produced vowels with a lower average f0 in the clear compared with casual speaking condition (see Figure 5), which suggests that they were aiming for more adultlike pronunciations of the words in question.

If children equate clear speech with adult speech, it is reasonable to wonder about the representations that they would access to speak in an adultlike manner. One possibility is that when planning speech, children access continuous articulatory representations that correspond to acquired lexical items. The lexical item may be stored as the composite of all input forms for that lexical item and may be weighted to reflect adult speech characteristics. Such a weighting assumes that language acquisition is motivated by a desire to communicate with others, and so, self-productions would have a minimal effect on the representation. Of course, the articulatory plan that a child develops to correspond with the lexical representation must have internal structure. This structure may be best described in discrete and paradigmatic terms (i.e., in distinctive feature or gestural terms), albeit those specific to the child's physiology and motor practice with speech. Nonetheless, if the objective is to speak like an adult, then the child's early clear speech strategy may be a strategy aimed at executing a whole word pattern that sounds roughly like an adult's rather than a strategy aimed at highlighting the internal structure of their own articulatory plan. If ever this plan were highlighted under more formal social conditions and undershot in casual ones, then the distinct styles could probably be characterized along a hyper–hypo speech continuum, just as in adult speech.

Acknowledgments

This work was supported, in part, by a Summer Research Award from The University of Oregon. We thank Tatiana Furrow, Jennifer Peddicord, and Gennifer Fink for help with participant recruitment and data collection. Gennifer Fink also contributed as the second rater. We are grateful to Robin High (formerly the University of Oregon statistical consultant and programmer) for developing sophisticated statistical models to accommodate the complex design of this study, Susan Guion for providing us with expert advice on the rating task and some aspects of the acoustic measures, Lou Moses for letting us use his lab to run subjects, and the University of Oregon Psychology Department for access to the baby and child database for participant recruitment.

References

Alexander, D., Wetherby, A., & Prizant, B. (1997). The emergence of repair strategies in infants and toddlers. *Seminars in Speech and Language, 18,* 197–213.

Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71,* 240–249.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.5.16) [Computer program]. Retrieved February 2007 from http://www.praat.org.

Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertent clear speech. *Speech Communication, 14,* 325–337.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech. I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 29,* 255–272.

Brinton, B., Fujiki, M., Loeb, D. F., & Winkler, E. (1986). Development of conversational repair strategies in response to requests for clarification. *Journal of Speech and Hearing Research, 29,* 75–81.

Cutler, A., & Butterfield, S. (1990). Durational cues to word boundaries in clear speech. *Speech Communication, 9,* 485–495.

Diehl, R. L., & Lindblom, B. (2002). Explaining the structure of feature and phonemic inventories. In S. Greenberg, W. Ainsworth, A. Popper, & R. Fay (Eds.), *Speech processing in the auditory system* (pp. 101–162). New York: Springer.

Ferguson, S., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 112,* 259–271.

Ferrier, S., Dunham, P., & Dunham, F. (2000). The confused robot: Two-year-olds' responses to breakdowns in conversation. *Social Development, 9,* 337–347.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Science, 10,* 244–249.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica, 38,* 148–158.

Genesee, F. (2001). Bilingual first language acquisition: Exploring the limits of the language faculty. *Annual Review of Applied Linguistics, 21,* 153–168.

Genesee, F., & Nicoladis, E. (1997). Language development in preschool bilingual children. *Journal of Speech-Language Pathology and Audiology, 21,* 258–270.

Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge, United Kingdom: Cambridge University Press.

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiological development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43,* 239–255.

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of lip and jaw control for speech. *Journal of Speech, Language, and Hearing Research, 45,* 66–79.

Hay, J. F., Sato, M., Coren, A. E., Moran, C. L., & Diehl, R. L. (2006). Enhanced contrast for vowels in utterance focus: A cross-language study. *The Journal of the Acoustical Society of America, 119,* 3022–3033.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect. *Language, 69,* 505–528.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *The Journal of the Acoustical Society of America, 77,* 266–280.

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America, 115,* 362–378.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lee, S., Potamianos, A., & Narayanan, H. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105,* 1455–1468.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America, 35,* 1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht, the Netherlands: Kluwer.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 565–604). Timonium, MD: York Press.

Lindblom, B., Brownlee, S., Davis, B., & Moon, S.-J. (1992). Speech transforms. *Speech Communication, 11,* 357–368.

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research, 44,* 670–684.

Moon, S.-J., & Lindblom, B. (1992). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America, 96,* 40–55.

Munhall, K. G., Kawato, M., & Vatikiotis-Bateson, E. (2000). Coarticulation and physical models of speech production. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 9–28). Cambridge, United Kingdom: Cambridge University Press.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal, 316,* 1236–1238.

Picheny, M. A., Durlach, N. L., & Braida, L. D. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96–103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434–446.

Riely, R. R., & Smith, A. (2003). Speech movements do not scale by orofacial structure size. *Journal of Applied Physiology, 94,* 2119–2126.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America, 118,* 1677–1688.

Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croation. *Journal of Phonetics, 36,* 91–113.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics, 13,* 335–349.

Storkel, H. L., & Morisette, M. L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools, 33,* 24–37.

van Son, R. J. J. H., & Pols, L. C. W. (1999). An acoustic description of consonant reduction. *Speech Communication, 28,* 125–140.

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research, 45,* 1119–1133.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language, 31,* 749–778.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3–29.

Received September 21, 2007

Revision received April 15, 2008

Accepted March 21, 2009

DOI: 10.1044/1092-4388(2009/07-0223)

Contact author: Melissa A. Redford, Department of Linguistics, University of Oregon, Eugene, OR 97403-1290. E-mail: redford@uoregon.edu.