Patterns of acquisition of native voice onset time in English-learning children

Joanna H. Lowenstein[a] and Susan Nittrouer[b]
Department of Speech and Hearing, Ohio State University, 110 Pressey Hall, 1070 Carmack Road, Columbus, Ohio 43210

(Received 2 November 2007; revised 16 May 2008; accepted 16 May 2008)

Learning to speak involves both mastering the requisite articulatory gestures of one's native language and learning to coordinate those gestures according to the rules of the language. Voice onset time (VOT) acquisition illustrates this point: The child must learn to produce the necessary upper vocal tract and laryngeal gestures and to coordinate them with very precise timing. This longitudinal study examined the acquisition of English VOT by audiotaping seven children at 2 month intervals from first words (around 15 months) to the appearance of three-word sentences (around 30 months) in spontaneous speech. Words with initial stops were excerpted, and (1) the numbers of words produced with intended voiced and voiceless initial stops were counted; (2) VOT was measured; and (3) within-child standard deviations of VOT were measured. Results showed that children (1) initially avoided saying words with voiceless initial stops, (2) initially did not delay the onset of the laryngeal adduction relative to the release of closure as long as adults do for voiceless stops, and (3) were more variable in VOT for voiceless than for voiced stops. Overall these results support a model of acquisition that focuses on the mastery of gestural coordination as opposed to the acquisition of segmental contrasts.
[DOI: 10.1121/1.2945118]

PACS number(s): 43.70.Ep, 43.70.Fq [CHS] Pages: 1180–1191

I. INTRODUCTION

How does children's speech acquire the phonetic structure of mature speakers in their linguistic environment? This question has intrigued child-language scientists for decades. Early accounts of language development were predicated on the notion that phonetic segments are integral to speech production; that is, they are the so-called "building blocks" of language. Accordingly, experimental approaches to language development were largely transcription based, focusing on when children learn to produce and combine discrete phonetic segments (e.g., Bean, 1932; Jakobson, 1941; Osgood, 1953; Prather, Hedrick, and Kern, 1975; Velten, 1943).

Some approaches to the study of speech development instead propose that the first unit of productive organization is something other than the phonetic segment, and these approaches use methods of analysis not focused as strongly on phonetic transcription. Generally, alternatives to the phoneme as the initial unit of speech organization focus on larger units, such as unanalyzed syllables or words. Some of the terminology invoked to convey this concept includes "word recipes" (Vihman and Velleman, 1989), "articulatory routines" (Menn, 1978, 1983), and "templates" (Vihman and Velleman, 2000). Speech development in this view involves gradually differentiating these routines into their component consonantal and vocalic gestures, and then learning to recombine these gestures as appropriate to produce sequences of speech with proper organization so as to allow the listener to retrieve phonetic structures (e.g., Goodell and Studdert-Kennedy, 1991). This process may be termed gestural reorganization.

A. Gestural reorganization

The notion that the organization of speech gestures is modified as children acquire experience producing and hearing their native language complements the concept of the phonetic segment as a perceptuomotor unit (e.g., Studdert-Kennedy, 1987). According to this account, phonetic segments do not exist separately from longer utterances. Rather, phonetic structure arises from the carefully coordinated set of gestures involved in the production of long stretches of speech. Children do not learn phonemes in this view, but rather discover phonetic structure by hearing how the gestures of their native language are produced and organized and simultaneously by learning how to accomplish that coordination themselves. In accordance with these models, some studies of speech development instead focus on analyzing patterns of articulatory movement in young children's speech through both acoustic analysis and direct measures of articulation (e.g., Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000).

In the speech of adults, individual gestures are precisely specified and generally produced with a great degree of consistency across utterances and exacting coordination among adjacent gestures such that phonetic implementation is in line with what is appropriate for the language being spoken and the situation in which an utterance is being produced [e.g., adaptive variability (Lindblom, 1990)]. It is this precision in spatial composition and interarticulatory timing rela-

[a]Electronic mail: lowenstein.6@osu.edu.
[b]Electronic mail: nittrouer.1@osu.edu.

tions that gives rise to segmental and suprasegmental structures in speech (e.g., Kelso *et al.*, 1984; Munhall, 1985; Nittrouer, 1991; Nittrouer *et al.*, 1988). However, unlike adults' speech, children's early speech gestures appear to be only roughly specified (i.e., spatial targets are general), to vary in form across similarly shaped utterances (which follows from them being roughly specified), and to lack precision in interarticulatory timing (Browman and Goldstein, 1989; Goffman and Smith, 1999; Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000; Macken and Barton, 1980a; Nittrouer, 1993; Nittrouer *et al.*, 1989; Nittrouer *et al.*, 1996; Smith and Goffman, 1998; Smith and McLean-Muse, 1986). Moreover, unlike the speech of adults, children's first babbling utterances show little language specificity. In fact, the patterns of early productions are strikingly similar across languages, at least when activity that can be correlated with specific segments is examined (e.g., Locke, 1983; Oller and Eilers, 1982).

However, when speech production is examined at a more "global" level, we find that children acquire the productive patterns of their native language at an early age. For example, de Boysson–Bardies *et al.*, (1986) derived the long-term spectra of adults with different native languages. When the long-term spectra of separate groups of 10-month-olds from those language environments were derived, it was found that those spectra differed across language groups in precisely the same way as the adults' spectra did. Thus, by 10 months of age children must know something about the gestural organization of their native language, although only at a global level, that is, one less clearly associated with individual phonetic segments and instead focused more on overall patterns of production.

A number of investigators have described the articulatory patterns of children's initial productions: specifically, babbling and/or first words (generally meaning when the child has less than roughly a dozen words). For example, MacNeilage and Davis (1991, 1993) reported that children's earliest utterances can be adequately described as cycles of mandibular elevation (resulting in poorly specified constrictions) and depression, accompanied by phonation; that is, children's early speech consists largely of complete closures of the lips or tongue against the palate (resulting in /b, d, m, n/) alternating with different degrees of mandibular opening (resulting in central vowels of varying heights). Stretches of long voicing lag are extremely rare in babbled productions (e.g., Locke, 1980; Whalen *et al.*, 2007), probably because young speakers are unable to incorporate vocal-fold abduction gestures into the already complicated (for young children) task of moving upper vocal-tract articulators while phonating.

Other studies suggest that at least some of young children's gestures actually differ from those of adults in form. For example, a motion tracking study of bilabial production found significant differences in articulator coordination between 1-year-olds, 2-year-olds, 6-year-olds, and adults (Green *et al.*, 2000). The 1- and 2-year-olds achieved oral closure primarily through jaw displacement, while the 6-year-olds and adults achieved oral closure through a combination of lower lip movement and jaw displacement. Furthermore, the contribution of the jaw to oral closure was found to differ significantly for 1-year-olds versus 2-year-olds, supporting the suggestion of reorganization at these early ages. In agreement with the findings of Green *et al.* (2000), Smith and McLean-Muse (1986) found that children as old as 11 years made labial gestures with different lower lip and jaw displacements and peak velocities than adults. Finally, coordination among gestures is not as precisely handled in children's speech as in the speech of adults: Goodell and Studdert-Kennedy (1993) recorded children (at 22 and 32 months) and adults imitating minimal pair nonsense disyllables that consisted of the syllable [bə] followed by a stressed [ba], [bi], [da], [di], [ga], or [gi]. They found that the proportions of total utterance duration assigned to each syllable became more adultlike at 32 months, and there was an age-related decrease in gestural overlap across the two syllables in terms of tongue front-back position. They also found evidence of more adultlike differentiation in tongue height for the schwa and stressed-syllable vowels at 32 months, compared with 22 months. Smith and Goffman (1998) found that children (aged 4 and 7 years) produced speech with less stable movement trajectories than adults (also see Goffman and Smith, 1999).

The process of gestural reorganization appears to continue into the early school years. For example, children as old as 7 years have been found to produce adjacent consonant and vowel gestures with more overlap than adults for both fricatives (Nittrouer *et al.*, 1989) and stops (Nittrouer, 1993); in addition, they have difficulty consistently coordinating vocal-tract closure with laryngeal abduction in the production of words with syllable-final voiced or voiceless stops (Nittrouer *et al.*, 2005). Several studies have found that children generally produce gestures at slower rates than adults, at least up to eight years of age (Goffman and Smith, 1999; Goodell and Studdert-Kennedy, 1993; Green *et al.*, 2000; Nittrouer, 1993; Smith and Goffman, 1998; Smith and McLean-Muse, 1986). Children have also been found to produce individual articulatory gestures with greater temporal variability than adults (e.g., Nittrouer, 1993). In sum, we find a much more convoluted and protracted developmental course for gestural speech reorganization than some phonetic transcriptions indicate.

B. Acquisition of syllable-initial stop voicing

The acquisition of voicing categories for syllable-initial stops can help inform us about developmental changes in gestural coordination. When stops occur at the start of words the timing between the release of the oral closure and the onset of phonation must be coordinated in order to provide the listener with appropriate information for voicing judgments. This timing, termed voice onset time (VOT), was first demonstrated to differ across languages by Lisker and Abramson (1964). Cross linguistically they found three patterns for VOT: long voicing lead, where phonation starts well before the oral release; short voicing lag, where phonation begins just after the oral release; and long voicing lag, where phonation begins well after the oral release. Because languages differ in their phonetic requirements, children need to learn the appropriate gestural timing specific to their native language.

Generally speaking, most words spoken are not produced in isolation; instead they occur in a stream of speech. In order to produce a voiceless aspirated stop, phonation is interrupted in continuous speech by a glottal abduction gesture. This gesture involves the posterior cricoarytenoid muscle, which begins to contract shortly before the glottis opens and reaches maximum activation before the peak of glottal opening. As the activity of the posterior cricoarytenoid muscle decreases, the interarytenoid muscles contract, resulting in the return of the glottis to a position appropriate for phonation (Löfqvist *et al.*, 1984; Löfqvist and Yoshioka, 1980). Because voiceless aspirated sequences are so rare in babbled speech (e.g., Locke, 1980; Whalen *et al.*, 2007), it is likely that children are unable to incorporate these abduction/adduction gestures into their productions. However, when they begin producing words, they must learn to do so.

Studies of stop acquisition using methods other than phonetic transcription have found that children only gradually develop the ability to produce long-lag VOTs. For example, Macken and Barton (1980a) audiotaped four English-learning children at 2 wk intervals starting around 18 months and ending around 26 months. Each taped session was 20–30 min in length, and toys were used in the sessions that would explicitly elicit stop productions. The examiners isolated words that began with stops, with the criterion that there had to be an interruption in voicing between the preceding word and the isolated target word. In other words, children in the study had at least mastered the ability to abduct or adduct vocal folds during continuous speech. Spectrographic and oscillographic analyses were done on the speech samples to measure VOT. Based on the results, Macken and Barton (1980a) proposed a three-stage model for the acquisition of stop consonants. These three stages are age independent and describe a linear path for the development of stop production. The first proposed stage of stop acquisition consisted of children producing primarily short-lag stops (defined as VOTs between 0 and 20 ms). In this stage, children's articulation seems to fit the description of "everything moves at once" offered by Kent (1983) because they appear to simultaneously, or nearly simultaneously, begin phonation and release the oral closure. Longitudinal studies of languages other than English confirm Macken and Barton's (1980a) findings that children initially produce primarily short-lag stops (Eilers *et al.*, 1984; Kehoe *et al.*, 2004; Macken and Barton, 1980b).

In Macken and Barton's (1980a) second stage, children begin to differentiate between word-initial voiced and voiceless stops. Words that have voiced initial stops in the adult version still have stops with VOTs close to 0 ms. However, in words that should have (according to the adult version) a voiceless aspirated initial stop, the stops are consistently produced with longer VOTs. Macken and Barton (1980a) suggested that this stage can be further divided into two substages. In the first, VOTs for stops intended as voiceless are consistently longer than those of stops intended as voiced but are clearly in the perceptual category of "voiced" for English listeners (i.e., less than 20 ms). In the second substage, these stops have VOTs on the order of 40 ms, and so are ambiguous in perceived voicing for adult English speakers. This second stage, where children are producing a contrast that is not clearly perceptible to adults [e.g., "covert contrast" (Scobbie *et al.*, 2000)], can be seen as evidence for gestural reorganization—children have begun to expand their options for coordinating voicing onset and oral release, but have not yet perfected the process.

Macken and Barton's (1980a) proposed third stage of VOT development is also described as having two substages. In the first, children produce voiceless stops with VOT means considerably longer than adult means (over 100 ms). In the second part of the third stage, children produce VOTs more similar to those produced by adults. In summary, Macken and Barton's (1980a) model of VOT acquisition is one suggesting that children initially have great difficulty producing long-lag VOTs; after some early modest attempts at delaying the onset of laryngeal vibration relative to closure release, children exaggerate this pattern, obtaining adultlike productions at just over 2 years of age.

Of course, the question arises as to why children do not produce the long-lag VOTs of their native language in their earliest stop productions: Do they fail to discriminate short- and long-lag VOTs in perception, or are they unable to coordinate laryngeal and supralaryngeal gestures appropriately? Numerous studies of infant speech perception have shown that infants as young as 2 months of age can discriminate syllables beginning with short-lag versus long-lag VOTs (e.g., Eimas *et al.*, 1971; Werker and Tees, 1999), and so we conclude that children's perceptual capacities are adequate to hear the differences in lag duration for English voiced and voiceless initial stops. Instead, the initial inabilities of infants to produce long-lag VOTs very likely represent production constraints.

Although elegant in design, Macken and Barton's (1980a) proposed model has not gone unchallenged. In particular, the suggestion that young children go through a period of producing voiceless VOTs longer than those of adults has mixed support in subsequent studies. Several studies have found that children up to six years of age produce shorter mean VOTs than adults for words intended to have initial voiceless stops (e.g., Kewley-Port and Preston, 1974; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976). Nittrouer (1993) found that VOT for initial /t/ produced by 3-year-olds is more than 10 ms shorter than those produced by older children and adults. There is also evidence that children attain the voiceless VOTs characteristic of their native language at varying ages: the 17 normal children in three related studies first produced adultlike VOT values at ages ranging from 18 to 29 months (Macken and Barton, 1980a; Snow, 1997; Tyler and Saxman, 1991).

C. Variability in VOT development

Regardless of claims about the patterns of VOT acquisition, a common finding of all studies is that children (at least up to 7 years of age) produce VOTs for voiceless stops with greater intraspeaker variability than adults (Barton and Macken, 1980; Gilbert, 1977; Kewley-Port and Preston, 1974; Koenig, 2000; Macken and Barton, 1980a; Menyuk and Klatt, 1975; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976). Variability is generally viewed by child phonologists as the reciprocal of skill in coordinating vocal-tract gestures; that is, variability diminishes as skill in coordinating gestures increases (e.g., Goodell and Studdert-Kennedy, 1993; Hodge, 1990; Kent, 1976). From this perspective, variability is viewed similarly to how investigators interested in the development of other skilled actions view it (e.g., Bruner, 1973), and diary studies are replete with reports of young children exhibiting great variability in their attempts at word productions. For example, Ferguson and Farwell (1975) listed ten transcribed attempts of a toddler attempting to say "pen," all of them different. It is clear from the list that the child had all the component gestures of the adult form, but was unable to coordinate timing among those gestures. In the case of VOT production, high variability relative to that which is encountered in more mature speakers likely signifies that the child is in the process of mastering the coordination between oral release and the onset of vocal-fold vibration that produces long-lag VOTs. We would not expect variability in VOT to be as great in children's attempts at voiced initial stops precisely because of the everything-moves-at-once principle: if one can begin all gestures at the same time, or nearly at the same time, coordination is simplified. For these reasons, the present investigation examined VOT variability, with the prediction that it would be greater for voiceless stops.

D. Avoiding difficult gestural sequences

One factor that complicates the study of VOT acquisition is that babbled productions and real words coexist in the speech of toddlers. Locke (1980) reviewed three studies of babbling in 11–12 month old American infants and found that complete closures of the vocal tract accompanied by even minimal vocal-fold abduction (i.e., voiceless aspirated stop-like sounds) were present in only 1–11% of babbled utterances; similarly, Oller and Eilers (1982) found that English- and Spanish-learning infants produced aspirated initial plosives only 6–11% of the time. This means that investigators are largely obliged to wait until children begin attempting words intended to have initial voiceless stops to examine the acquisition of these gestural sequences. However, what if, as some have suggested (Schwartz and Leonard, 1982; Schwartz *et al.*, 1987), children avoid words with gestural sequences that are difficult for them? This could complicate analyses, particularly those that rely on counts of voiceless and voiced syllable-initial stops. For example, Kewley-Port and Preston (1974), who analyzed all vocalizations that began with non-nasal, nonlateral, pulmonic egressive apical stop consonants (both babble and words), found for two of the children in their study that there was an increase in the number of apical stops with long-lag VOTs (defined as over 25 ms) from less than 15% of all apical stops to more than 50% between the ages of 1 and 2 years. It is possible that this change resulted from an increase in the number of words (rather than babbled sequences) with intended voiceless stops, although it is not clear from their report because they did not distinguish between the words and babble. In a study of spontaneous speech, Dobrich and Scarborough (1992) found that children between the ages of 2 and 3 years produced fewer target words, which in their adult form contained final consonant clusters, than their mothers did, showing some evidence of avoidance for children in that age range.

However, the avoidance hypothesis begs the question of how children acquire production skills: If they are truly avoiding words with certain gestural sequences, how do they ever get the practice needed to master those sequences? Schwartz *et al.* (1987) answered that question by suggesting that children analyze the segmental contents of words in order to determine whether they are currently able to produce them and gradually relax their selection constraints in what they will attempt as they gain more general speaking experience. We might modify the account to suggest that children analyze the *gestural* demands of words and avoid those for which those demands are too great. Then, as they gain more experience generally producing speech, they become more willing to try new gestural sequences. In any event, counting the numbers of words produced by children that in the adult form begin with long-lag VOTs should reveal whether children initially avoid words that begin with voiceless aspirated stops, and then increase their attempted productions of these words as they become more experienced.

E. The current study

The present investigation explored developmental trends in gestural coordination by focusing on the acquisition of VOT because long-lag voiceless stops require careful timing between the oral release and the onset of phonation, are rarely found in babble and early words, and only slowly approach adult VOT values. This protracted developmental course of voiceless initial stop acquisition makes this gestural pattern an ideal candidate for study. The investigation posed three specific questions: First, do children's productions of VOT for syllable-initial stops intended to be voiceless become more like those present in their native language by means of a gradual acquisition process? The answer to this question would inform us regarding how children's speech productions acquire phonetic structure. Finding that children only gradually master the coordination needed for initial stops to be heard as voiceless aspirated stops would support the position that phonetic structure emerges as children master component speech gestures and language-specific coordination among them. The specific hypothesis to be tested stemmed from Macken and Barton's (1980a) three-stage model, but with modifications: children less than three years of age were expected to (1) initially produce only short-lag VOTs for stops intended as both voiced and voiceless, then (2) produce a difference in VOTs for voiced and voiceless stops that do not correspond to adult values, and (3) finally produce VOTs that approach but are still shorter than adult values. The prediction for the third stage of this model differs from that of Macken and Barton (1980a) but is commensurate with the results of others (Kewley-Port and

Preston, 1974; Nittrouer, 1993; Zlatin and Koenigsknecht, 1976), showing that children continue to have shorter VOTs than adults for voiceless initial stops into the early school years.

The second question asked in the present study was whether children become more skilled at achieving voiceless VOT targets during the ages studied here. Measuring variability for each speaker individually provides an estimate of how consistent speakers are in their productions. We predicted that variability would be greater in children's productions of words with voiceless initial stops, rather than with voiced initial stops, but we did not know if that variability would diminish over the course of the study.

The third question asked was whether children increase the number of stops intended as voiceless as they gain experience in producing their native language. The specific hypothesis to be tested was that there would be a steady increase in the number of stops intended as voiceless over time, demonstrating that children are willing to produce more stops intended to be voiceless as they become more experienced. In other words, evidence of this finding would support the notion of avoidance.

The speech samples that were analyzed for this study were obtained from children in unstructured play situations with one of their parents. The speech samples used in these analyses were collected with the purpose of tracking the emergence of gestural organization for many kinds of phonetic sequences in young children as they acquire their first words. Alternative methods of collecting speech samples include imitation and targeting specific utterances, for example, through the use of toy animals that are given names with desired phonetic sequences. While each of these methods is appropriate for certain research goals, the purpose in our method was to obtain a naturalistic sample of what actually happens during that early word acquisition period.

II. METHOD

A. Participants

Participants were all part of a larger study (McGowan *et al.*, 2004) in which seven children (three males and four females) were each followed for almost a year and a half. Nine children were originally recruited, but two left the study before all recording sessions could be completed. The children recruited were typically developing children. All had normal prenatal histories, normal deliveries, and no special medical conditions. None of the children had an immediate family member with speech, language, or hearing disorders. None of the children had any reported history of otitis media with effusion at the start of the study, and no child was treated for more than one episode while the study was being conducted.

Children were taped at 2 month intervals between roughly the ages of 14 and 31 months. Recording sessions began when the parents estimated that the child had ten recognizable words, and were discontinued when the child began stringing three or more words together to form sentences. Children were not all recorded at precisely the same ages because they reached the landmark point of having ten identifiable words at slightly different times. By beginning recordings at the same point in language development, we hoped to obtain samples across children at consistent "language ages," so to speak. For these analyses, children's data were grouped into 2 month intervals across the study (15–16, 17–18, 21–22, 23–24, and 27–28 months), with a few exceptions. Not all children could be recorded at the ages of 19–20 or 25–26 months because of illness or family vacations. Not all children were recorded at 30–31 months because some had passed the three-word sentence criterion for dismissal from the project. Consequently, these three intervals were not included in the current study. Nonetheless, the sessions for which data were analyzed can provide a good overview of developmental trends. Within these intervals, children were judged to be at approximately the same developmental age, as judged by vocabulary size and general utterance length.

B. Equipment and materials

Recordings were obtained using an AKG C-535EB microphone, a Shure model M268 mixer, and a Nakamichi MR-2 cassette deck with metal tapes. This system had a flat frequency response out to 20 kHz. These recordings were subsequently low-pass filtered with a high-frequency cutoff of 10 kHz and digitized with a Soundblaster A/D card using the SPEECH STATION II software at a 22.05 kHz sampling rate and 16 bit digitization.

The same set of toys was available for play for all children at all sessions and consisted of some small stuffed dolls, foam puzzles, and cloth books. All toys used in these sessions were soft to minimize extraneous noises that might interfere with speech recording. The toys were not chosen to elicit specific responses from the children but were rather meant to stimulate communication in general. However, there were toys present that were balanced in terms of the voicing category of the initial stop in their referent lexical items, such as a foam puzzle of a pig (/p/) and a doll of Bert from Sesame Street (/b/).

C. Procedures

Children were recorded in the same sound-proof booth at each session. Our goal was to have 20 min sessions, but sometimes a turn for the worse in the child's behavior forced us to curtail a recording session. However, all sessions were at least 11 min in length. The child sat on a high chair at a table, with one parent across the table. The microphone was suspended roughly 23 cm above the child's mouth. It was suspended rather than table mounted because pilot work showed that children habituated to its presence more rapidly that way. Parents were instructed to play with their children, trying to elicit as much language as possible. Between sessions parents were asked to keep diaries of new vocabulary items (at the younger ages) and new sentence structures (at the older ages) that they heard at home.

D. Measurements

For this study, 10 min worth of speech was generally analyzed. 10 min sections were obtained by finding the temporal middle of each recording session and using 5 min on either side of that point. If the original session length was only 11 min, the entire session was analyzed. For each child, the recordings were analyzed in "backward" order, starting with the one obtained at the oldest age, then the one at the next-to-oldest age, and so on until all recordings were analyzed. This allowed the examiner to acclimate to each child's speech, starting with what should have been the most intelligible sample.

Words heard as starting with an initial stop were extracted and yielded a total sample of 1458 words. Words that closely matched the adult form and those that did not were included. Words that were not in phrase-initial position were included only when there was a clear break in voicing after the end of the previous word. Words were excluded if there was voicing carried through the closure, as it is never clear what portion of that voicing belongs to the closure of the previous word and what portion belongs to the target word (e.g., Allen, 1985). Furthermore, voicing during oral closures of considerable duration in children's speech is often a consequence of nasalization. As such, that voicing could not be counted as "prevoicing." The first author made phonetic transcriptions of each word, and a phonological transcription according to what word was intended. Whenever possible, these transcriptions were based on a conversational context (e.g., parental repetition of the word, what book the child was looking at, or what toy the child was playing with). For example, [bɪg] might be transcribed as "big," or "pig," depending on the context: If the parent and child were discussing something that was big, it was transcribed as "big;" if they were discussing farm animals, it was transcribed as "pig." The context was ambiguous for 24% of words (350), across all sessions. For these instances when conversational context was ambiguous, transcriptions were based on what the transcribers heard and thought the child intended. When context did not clearly signal word identity, two research assistants also transcribed the word. Generally there was little disagreement about the intended voicing of the stop among the three transcribers; slightly more were transcribed as voiceless than as voiced. (Place was more problematic, but VOT according to place was not considered here.) Specifically, the first author and research assistants identified 90% of the words (315) with ambiguous context as starting with the same voicing target. In the 10% of words for which there was disagreement (35), words were transcribed according to the decisions of the first author, who had the most experience listening to the individual children and so knew what their interests and speech patterns generally were. In particular, the first author had the benefit of being familiar with what were actually *later* produced words from the child because of the backward analysis approach. This sort of ambiguity in word identity is unavoidable when speech samples are collected from unstructured play situations, but again, the availability of samples such as these is important to our collective understanding of phonetic development.

Three measurements were made on words from each session for each child. First, the numbers of words with voiced and voiceless initial stops in the 10 min section were counted. Stops were counted as voiced or voiceless depending on what they would be in the adult version of the word. Second, VOT was measured for each stop from a combined wave form and spectrogram using WAVESURFER (Sjölander and Beskow, 2000). Markers were placed at the start of the broadband aperiodic burst in the wave form (correlate of the oral release) and at the onset of a regular periodic signal in the wave form (correlate of voicing onset), as illustrated in Fig. 1. Marker placement was confirmed in the spectrographic display. VOT was computed as the interval between these two markers. VOTs with more than 50 ms of prevoicing were excluded because of the rarity of such long-lead voicing in initial position in English. Moreover, for these particular samples, it was usually clear that these long-lead segments were not true oral vocalizations, but rather nasal productions. This observation is what would be expected for children, given their small oral cavities (which would make it hard to continue airflow through the glottis for very long before sub- and supraglottal pressures equalized, causing vocalization to cease). Sequences with more than a 300 ms lag were also excluded. Gaps that are long between an oral release and a vocalic segment exceeded the average syllable length for most of these children, and so it was reasonable to judge that the two elements (release and vocalic production) could not be viewed as part of a single integrated syllable. For each child, mean VOT was calculated for voiceless and voiced target initial stops separately.

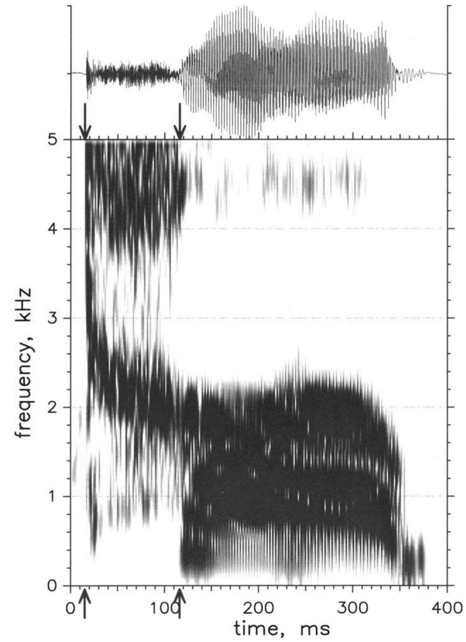

FIG. 1. Wave form and spectrogram of a 24-month-old male's production of /tub/. Arrows indicate VOT measurement boundaries.

As the avoidance hypothesis would suggest, some children within some recording sessions did not produce many words with particular initial stops, making it hard to obtain a sample large enough to provide an accurate estimate of VOT for that particular stop. Therefore, if fewer than five tokens of a specific stop (/b/, /d/, /g/, /p/, /t/, or /k/) were present in the 10 min segment identified for use, VOT for that stop was analyzed from the entire recording session. This was done in order to obtain more reliable estimates of VOT across place.

TABLE I. Number of children (out of seven) who produced fewer than five stops in a category during the middle 10 min of recording, by session.

Age (month)	Stop					
	/b/	/d/	/g/	/p/	/t/	/k/
15–16	4	4	6	7	7	7
17–18	3	2	6	5	7	7
21–22	0	3	4	7	5	5
23–24	0	2	6	5	4	3
27–28	0	1	4	3	1	1

Table I lists the numbers of children (out of seven) who produced fewer than five stops in each category during the 10 min segment at each age. These additional tokens resulted in total sample sizes of 140 (15–16 months), 209 (17–18 months), 325 (21–22 months), 364 (23–24 months), and 420 (27–28 months) being used for the computation of VOTs.

The third measurement made was variability for each child's mean VOT for voiced and voiceless target initial stops separately. Standard deviations (SDs) were used to index variability. Coefficients of variation, computed by dividing SD by the mean as a way to normalize SD, are sometimes used to index variability in temporal speech measures (e.g., Smith, *et al.* 1983). However, that method is generally reserved only for situations in which measures have an absolute limit of 0 ms (such as syllable length). Because initial stops can be prevoiced, that situation does not exist in this case. Within-child standard deviations (WCSDs) were calculated for each child's voiced and voiceless initial stops.

Statistical analyses were performed using the BMDP statistical software (1990). The magnitude of the voicing effect on measures was computed using both η^2 and Cohen's d. η^2 varies between 0 and 1 and measures the proportion of variance accounted for by a single factor. Cohen's d measures the relative magnitude of difference between two means in terms of standard deviation and is therefore a normalized index of effect size (Cohen, 1988). Generally speaking, $d > 0.8$ represents a large effect size.

III. RESULTS

A. Description of children's productions

At 15–16 months, all seven children were producing primarily one- or two-syllable utterances, with one or two identifiable words produced in a session. At 17–18 months, two of the children were mostly producing identifiable words, while the other five were producing combinations of babbled sequences and real words. At 21–22 months, three children were producing only words, and four were producing mostly words. At 23–24 months, four children were producing only words, and three were producing mostly words, generally in a mix of one-word and two-or-more-word utterances. At 27–28 months, all of the children were producing only identifiable words, generally in two-or-more-word utterances. According to parent reports, all of the children in this study were producing ten words at their first recording session (15–16 months). The children reached the point where they were producing 50 words at sometime between 18 and 26 months (mean 22.6 months, median 24 months).

B. VOT

Figure 2 presents VOT averaged across all children and voicing categories (i.e., all tokens from all children were averaged for each time period), with error bars representing standard deviation across all tokens from all children. The number of tokens included in the calculations at each age is indicated at the bottom of the graph. This analysis provides a picture of the overall changes in VOT regardless of voicing category assignment. Mean VOT was 12 ms at the youngest age examined and increased to 40 ms at 23–24 months, where it remained. Standard deviations for VOT also increased over time. Looking at the standard deviation, it is clear that VOTs at 15–16 months fell into a relatively narrow range (−10 to+30 ms) that roughly corresponds to the voiced category in English. This range also includes prevoiced stops, which are relatively uncommon in the initial position in American English. At 23–24 and 27–28 months, children's VOTs span the range of values for adult speakers of American English. A one-way analysis of variance (ANOVA), with age as the categorical variable, was performed on VOT, revealing a significant difference in VOT across age groups: $F(4, 1453)=31.475$, $p<0.001$, $\eta^2=0.80$. Thus, age accounted for a large proportion of variance in VOT (80%).

Figure 3 presents mean VOT for stops judged to be intended as voiced and voiceless, separately. To investigate developmental changes in VOT for voiced and voiceless stops separately, a simple effect analysis was done, investigating age-related changes in VOT for each voicing category

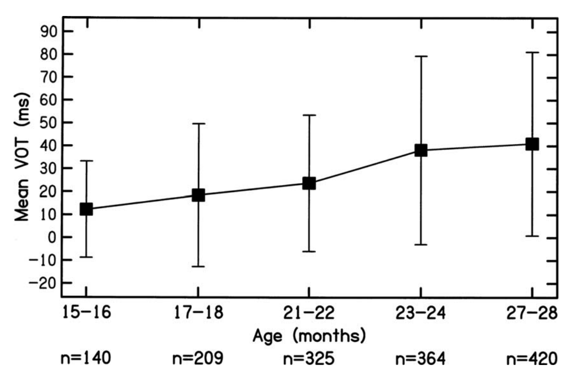

FIG. 2. Mean VOT values for children's stop productions, by age. Error bars represent standard deviations. Total sample sizes for each age are indicated.

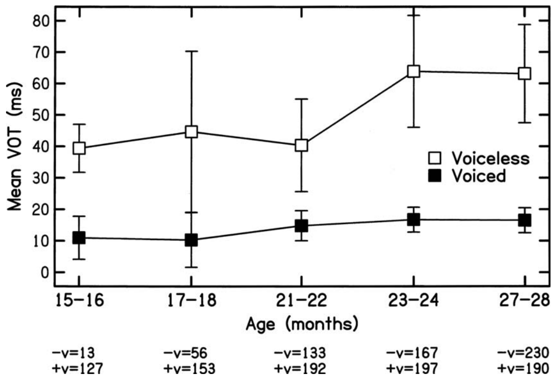

FIG. 3. Mean VOT values for children's voiced and voiceless stop productions, by age. Error bars represent standard deviations. Total sample sizes for voiced and voiceless stops for each age are indicated.

(i.e., voicing was held constant). A simple effect analysis is an appropriate selection of statistical tests for experiments with several independent factors because it permits the examination of effects for one or more of those factors at each level of another factor while using the overall estimate of error variance. This procedure provides a more sensitive test than would be obtained by doing separate ANOVAs at each level of the factor. Significant results were obtained for both the voiced [$F(4,24)=2.88$, $p=0.045$] and voiceless [$F(4, 24)=4.26$, $p<0.01$] stops. Children's mean voiced VOTs started at 11 ms, and stabilized at 16 ms at 21–22 months. Mean voiceless VOTs were constant at roughly 40 ms through 21–22 months. Then, at 23–24 months, the mean VOT jumped to above 60 ms. The relative magnitudes of the age differences on VOT can be indexed by computing the difference in mean VOT at each of these times and dividing by the pooled standard deviation (Cohen's d). Comparing children's initial voiced VOTs to their stable voiced VOTs (at 21–22 months) gives a d of 0.65 (a medium effect size), while comparing children's initial voiceless VOTs to their stable voiceless VOTs (at 23–24 months) gives a d of 1.79 (a large effect size). The error bars on Fig. 3 indicate standard deviation across children. It is clear that voiceless stops were produced with more variability in VOT (greater standard deviations) than voiced stops. It is also interesting that standard deviations for voiced stops decreased over the course of the study, while standard deviations for voiceless stops increased after the first recording session and remained high. It appears that these children were producing voiced stops more accurately with experience (resulting in lower variability) but were not yet able to perfect the coordination of gestures for producing voiceless aspirated stops (resulting in higher variability). These standard deviations were larger than those found for both adults and older children in previous studies (Koenig, 2000; Nittrouer, 1993).

C. Within-child standard deviations (WCSDs)

Very few syllable-initial stops meeting the criterion of "voiceless" were found in the samples of 15–16 month olds. With so few tokens from each child, it would not be appropriate to compute WCSDs. Consequently, statistics for

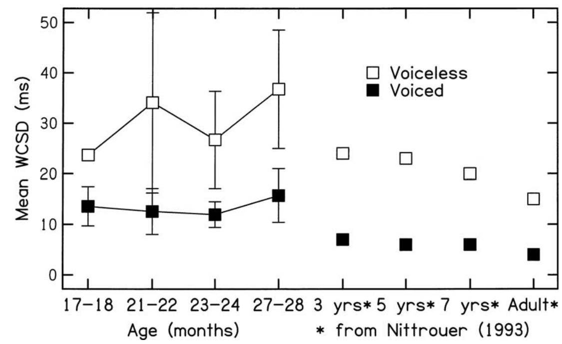

FIG. 4. Mean VOT WCSD, by age, including data from Nittrouer (1993). Error bars represent standard deviations.

WCSD were calculated beginning with the 17–18 month session. The left half of Fig. 4 displays mean WCSDs for each intended voicing category, across recording sessions, with error bars representing WCSD standard deviations. It appears that variability in VOTs for children's voiced initial stops was low and relatively consistent, with WCSDs of 12–16 ms. Variability VOTs for their voiceless initial stops, on the other hand, was greater, with mean WCSDs of 24–37 ms. A factorial analysis was performed on WCSD, examining the effects of age and voicing. Only the main effect of voicing was found to be significant, $F(1, 6) = 44.23$, $p<0.001$, $\eta^2=0.82$, supporting the conclusion that WCSDs were greater for the voiceless than for the voiced stops. WCSD did not change significantly over time, for either the voiced or voiceless stops. As a comparison, note that Koenig (2000) reported mean within-subject standard deviations for voiceless stops of 11 ms for adults and 21 ms for 5-year-old children, and both of these standard deviations are shorter than those found for voiceless stops for our subjects at any of the ages we studied. We conclude that at least over the age range examined in this study, children did not improve in their abilities to coordinate the vocal-tract release and the onset of voicing.

D. Counts of words with initial voiceless and voiced stops

Figure 5 displays mean numbers of voiced and voiceless target stops, as counted from each 10 min segment; that is, stops outside of this time window that were examined to help provide an estimate of VOTs were not included in this analy-

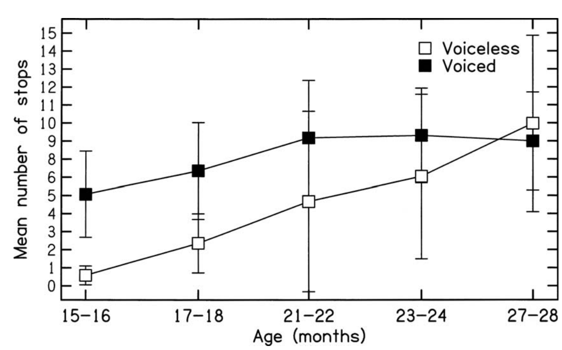

FIG. 5. Mean number of voiced and voiceless stops, by age. Error bars represent standard deviations.

sis. In examining Fig. 5, it is clear that the numbers of target voiced stops produced remained relatively constant over time, while the growth in target voiceless stops was nearly linear. Children started out attempting less than one voiceless stop on average during the 10 min analysis segment at 15–16 months and were producing similar numbers of voiced and voiceless stops by the end of this series of recordings.

To examine the developmental changes in stop count for voiced and voiceless stops separately, a simple effects analysis was done, investigating age-related changes in count for each target voicing category. There was no significant age effect for the numbers of voiced stops, but there was for the numbers of voiceless stops, $F(4,24)=6.55$, $p=0.001$. This effect reflects the steady growth in the numbers of words with intended initial voiceless stops produced by children across recording sessions. The relative magnitude of the change in the count of voiceless stops between 15–16 and 27–28 months was calculated using Cohen's d, resulting in a d of 2.42. This is clearly a very large effect.

E. Individual developmental patterns

Group averages inform us about general developmental trends across children, but cannot tell us if all children follow a similar developmental pattern. The most important information that longitudinal studies such as this one provides concerns patterns for individual children. It may be that the group trends reported here do not hold for all children. Perhaps one or a few of the children began exhibiting adultlike VOT patterns as soon as they began producing their first words. In fact, when individual patterns of development were examined, it was found that no child in this study exhibited an adultlike behavior in VOT production from the start of first words. In particular, all of the children produced very few target initial stops that could be considered to be intended as voiceless during the earliest two test sessions (note, however, that six out of seven children produced at least one). In other words, there was no evidence that the apparent avoidance of word-initial voiceless stops was characteristic only of a subset of the children. This finding was interesting because many of the items that children this age are typically interested in and, indeed, that these particular children talked about at later ages, begin with voiceless stops, for example, cats and cars. Again, the procedures of data collection for this study did not specifically encourage or dissuade children from using particular words, but the opportunities were present for producing words with both voiced and voiceless stops.

Examining VOTs of individual children showed two distinct patterns of acquisition. Four of the seven children in this study followed the overall pattern described earlier: All VOT measures were within the adult category of voiced during the earliest session, and then VOTs for words with intended initial voiceless stops increased to over 60 ms during the course of the study. The other three children showed a slightly different pattern of acquisition. For these children, VOT measures for words with intended voiceless stops were initially on the low end of the adult range; that is, VOTs were within 40–55 ms for these children. Over the course of the study, two of these children showed increases in VOTs for words intended to have initial voiceless stops. One of these children, however, never showed any change in VOTs for words intended to have initial voiceless stops, even though she did increase the number of words in this category that she attempted. Finally, all children in the study showed the same pattern regarding WCSD: when VOT for words intended to have voiceless initial stops was on the low end of the adult range (i.e., within 40–60 ms) or adultlike (i.e., over 60 ms), variability was high.

In summary, whether or not these children started out with clearly short-lag or ambiguous VOTs for voiceless stops, they all went through some sort of a developmental process. For most of the children, this process involved changes in the number of initial stops intended as voiceless that they attempted and increases in VOT itself. For one child, however, this developmental process demonstrated itself as only a change in the number of stops intended as voiceless. This child never achieved adultlike VOTs for words intended to start with voiceless stops. All children showed a lack of skill in coordinating supralaryngeal and laryngeal gestures for the production of voiceless VOT, as indicated by the finding of great variability through the entire length of this study.

IV. DISCUSSION

The goal of this study was to examine developmental trends in gestural coordination for VOT. We posed three specific questions: The first question was whether children's VOT for syllable-initial stops intended to be voiceless become more like those present in their native language later than VOT for syllable-initial stops intended to be voiced. It is clear that the children in this study followed a gradual acquisition process in their productions of voiceless VOTs. Looking at overall patterns (Fig. 3), mean voiceless VOTs remained in the short-lag or ambiguous range (i.e., below 55 ms) between 15–16 and 21–22 months. At 23–24 months, mean VOT increased so that it was in a range where stops would consistently be categorized as voiceless by native English speakers (i.e., above 60 ms). Looking at individual results, four of the seven children in this study performed according to Macken and Barton's three-stage model (1980a): voiceless stops were produced with short-lag VOTs until 21–22 months; then voiceless stops were produced with ambiguous VOTs; and at 23–24 months, voiceless stops were produced with VOTs closer to mature values. Two of the other three children achieved more mature VOTs at an earlier age, but still started out at 15–16 months producing VOTs in the ambiguous range. The seventh child did not acquire mature voiceless VOTs during the course of this study, even though values were within the ambiguous range from the onset of first words. According to these collective findings, we conclude that children only gradually modify their target voiceless VOTs to be similar to those of adults in their language community.

The second question asked was whether children would become more skilled at achieving VOT targets over the course of this study, as would be indicated by significant reductions in within-speaker VOT variability (i.e., WCSD). Results showed that WCSDs stayed fairly constant during the time course of this study, with voiceless stops showing more variability than the voiced stops (Fig. 4). Thus, variability in the organization of laryngeal and supralaryngeal gestures for the production of long-lag VOT does not diminish until older ages than those we examined in this study. So, while children's target VOTs for voiceless initial stops may have become more adultlike over the course of this study, their skill at achieving those targets did not improve (taking variability as the reciprocal of skill, e.g., Goodell and Studdert-Kennedy, 1993; Hodge, 1990; Kent, 1976). The right half of Fig. 4 presents variability data for children and adults from Nittrouer (1993). This graph shows that variability in the coordination of upper vocal tract and laryngeal gestures decreases with increased experience but that this decrease is not seen until children are older than they were when recordings were made for this study. This result for the coordination of laryngeal and supralaryngeal gestures in the production of *syllable-initial* stops labeled as voiced and voiceless mirrors what was found for the coordination of gestures in the production of *syllable-final* stops labeled as voiced and voiceless by Nittrouer *et al.* (2005): in that study, the duration of the vocalic syllable portion preceding voiced and voiceless final stops was measured. It was found that vocalic duration was more variable in samples from 5- and 7-year-olds than in those from adults, and this age-related trend was found only for syllables with voiceless final stops. As with syllable-initial voiceless stops, the timing of a supralaryngeal gesture (in this case, closing) must be precisely coordinated with a laryngeal gesture (in this case, abduction) in the production of syllable-final voiceless stops. The coordination of these gestures is neither as critical nor as difficult for syllable-final voiced stops as for syllable-final voiceless stops: final stops are categorized as voiced as long as voicing stops after the vocal tract closes, and that happens naturally when the pressure in the closed space of the supraglottal cavity equals the pressure of the subglottal cavity. For voiceless final stops, the speaker must make an explicit abduction gesture at just the right time.

The third specific question asked by us was whether children increase the number of stops intended as voiceless as they gain experience in producing their native language. The highly significant near-linear growth in the count of stops intended as voiceless (Fig. 5) indicates that these children did. Our procedures were sufficiently unstructured to have allowed children to produce as many words with either voiced or voiceless stops as they wanted. Nonetheless they showed a preponderance of words with voiced initial stops, particularly before 23–24 months. This evidence supports the position of others that children may avoid using words that they have difficulty producing during the earliest stage of word production (e.g., Schwartz and Leonard, 1982).

This study examined articulatory gestures in early speech production, focusing on the acquisition of stop consonants because voiceless stops require careful coordination in timing between the release of the oral closure and the onset of phonation. We found evidence that children gradually develop the timing between the larynx and the upper vocal-tract articulators that permit the perception by others of voiceless stops, supporting the concept of gestural reorganization over the first couple of years of life. The findings reported here also speak of the desirability of studying acquisition through the lens of gestural coordination: as children gained language experience, they were found to shift the timing between laryngeal and supralaryngeal gestures as needed to support the perception of voiceless initial stops. If we had not examined interarticulator coordination, all we could have reported would have been that voiced initial stops dominate early productions. There was also evidence of differences in developmental patterns among individual children in this study, but in no case did a child produce perfectly articulated and timed voiceless VOTs from the onset of first words.

The use of speech samples collected in unstructured play settings resulted in some challenges, but also strengthened the conclusions of this study and those of other investigators, as well. The biggest challenge was that decisions about whether specific stops were intended to be voiced or voiceless were not as clear cut as if an imitation task or a task with prenamed stimulus materials had been used. On the other hand, information was gathered about what actually happens in a naturalistic setting. Until now, the avoidance hypothesis has been tested primarily by presenting prenamed stimulus materials to young children and having them speak those names. Therefore, it was not known if children under the age of 2 years actually avoided the use of words with difficult (for them) phonetic sequences in natural settings. The results of this study show that children do appear to engage in such avoidance. That is, in typical conversational exchanges with a parent where no specific target was being solicited, children produced words with less complex articulatory coordination more frequently than those requiring more complex articulatory coordination.

In summary, this study found evidence to support the position that phonetic structure gradually emerges in children's speech production as they gain experience with their native language. The first words of four of these seven children adhered to the everything-moves-at-once principle in that the larynx was adducted at the same time that the upper vocal tract moved away from stop closure. Three of the children were able to delay the onset of voicing for a bit, but their laryngeal and supralaryngeal gestures were clearly not organized the same as those of adult speakers in their language community. All but one of the children showed evidence of reorganizing their speech-related gestures over the course of this study, but their variability remained high, indicating that their consistency in articulatory organization remained poor. Finally, all children initially seemed to avoid words that required the difficult task of delaying the onset of voicing relative to vocal-tract opening.

ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC00633 from the National Institute on Deafness and Other

Communication Disorders, the National Institutes of Health, to S.N. We thank Carol Manning and Gina Meyer for their assistance in data collection and digitizing. Portions of this work were presented at the 147th Meeting of the Acoustical Society of America, New York, May 2004.

Allen, G. D. (**1985**). "How the young French child avoids the pre-voicing problem for word-initial voiced stops," J. Child Lang **12**, 37–46.

Barton, D., and Macken, M. A. (**1980**). "An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children," Lang Speech **23**, 159–169.

Bean, C. H. (**1932**). "An unusual opportunity to investigate the psychology of language," J. Genet. Psychol. **40**, 181–202.

BMDP Statistical Software (**1990**). BMDP Statistical Software, Inc., Los Angeles.

Browman, C. P., and Goldstein, L. (**1989**). "Articulatory gestures as phonological units," Phonology **6**, 201–251.

Bruner, J. S. (**1973**). "Organization of early skilled action," Child Dev. **44**, 1–11.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum, Hillsdale, NJ).

de Boysson–Bardies, B., Sagart, L., Halle, P., and Durand, C. (**1986**). "Acoustic investigation of crosslinguistic variability in babbling," in *Precursors of Early Speech*, edited by B. Lindblom and R. Zetterström (Stockton, New York), pp. 113–126.

Dobrich, W., and Scarborough, H. S. (**1992**). "Phonological characteristics of words young children try to say," J. Child Lang **19**, 597–616.

Eilers, R., Oller, D. K., and Benito-Garcia, C. R. (**1984**). "The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study," J. Child Lang **11**, 313–336.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (**1971**). "Speech perception in infants," Science **171**, 303–306.

Ferguson, C. A., and Farwell, C. B. (**1975**). "Words and sounds in early language acquisition," Language **51**, 419–439.

Gilbert, J. H. (**1977**). "A voice onset time analysis of apical stop production in 3-year-olds," J. Child Lang **4**, 103–110.

Goffman, L., and Smith, A. (**1999**). "Development and phonetic differentiation of speech movement patterns," J. Exp. Psychol. Hum. Percept. Perform. **25**, 649–660.

Goodell, E. W., and Studdert-Kennedy, M. (**1991**). "Articulatory organization of early words: From syllable to phoneme," in Proceedings of the XIIth International Congress of Phonetic Sciences (Aix-en-Provence), pp. 166–169.

Goodell, E. W., and Studdert-Kennedy, M. (**1993**). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study," J. Speech Hear. Res. **36**, 707–727.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). "The physiologic development of speech motor coordination: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**, 239–255.

Hodge, M. M. (**1990**). "Measuring coarticulation in children's speech: Quirks and questions," Paper presented at the Child Phonology Conference, Madison, WI, May.

Jakobson, R. (**1941**). *Kindersprache, Aphasie und Allgemeine Lautgesetze*. Reprinted as *Child Language, Aphasia, and Phonological Universals*, 1968. Mouton, The Hague.

Kehoe, M. M., Lleo, C., and Rakow, M. (**2004**). "Voice onset time in bilingual German-Spanish children," Bilingualism: Lang. Cognit. **7**, 71–88.

Kelso, J. A., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (**1984**). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," J. Exp. Psychol. Hum. Percept. Perform. **10**, 812–832.

Kent, R. D. (**1976**). "Tutorial: Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies," J. Speech Hear. Res. **19**, 421–445.

Kent, R. D. (**1983**). "The segmental organization of speech," in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 57–89.

Kewley-Port, D., and Preston, M. S. (**1974**). "Early apical stop production: A voice onset time analysis," J. Phonetics **2**, 195–210.

Koenig, L. L. (**2000**). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," J. Speech Lang. Hear. Res. **43**, 1211–1228.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by J. W. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 403–439.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Locke, J. L. (**1980**). "Mechanisms of phonological development in children: Maintenance, learning, and loss," in *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society* (Chicago Linguistic Society, Chicago), pp. 220–238.

Locke, J. L. (**1983**). *Phonological Acquisition and Change* (Academic, New York).

Löfqvist, A., McGarr, N. S., and Honda, K. (**1984**). "Laryngeal muscles and articulatory control," J. Acoust. Soc. Am. **76**, 951–954.

Löfqvist, A., and Yoshioka, H. (**1980**). "Laryngeal activity in Swedish obstruent clusters," J. Acoust. Soc. Am. **68**, 792–801.

Macken, M. A., and Barton, D. (**1980a**). "The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants," J. Child Lang **7**, 41–74.

Macken, M. A., and Barton, D. (**1980b**). "The acquisition of the voicing contrast in Spanish: a phonetic and phonological study of word-initial stop consonants," J. Child Lang **7**, 433–478.

MacNeilage, P. F., and Davis, B. (**1991**). "Acquisition of speech production: Frames, then content," in *Attention and Performance XIII*, edited by M. Jeannerod (Erlbaum, New York), pp. 453–476.

MacNeilage, P. F., and Davis, B. L. (**1993**). "Motor explanations of babbling and early speech patterns," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Kluwer Academic, Dordrecht), pp. 341–352.

McGowan, R. S., Nittrouer, S., and Manning, C. J. (**2004**). "Development of [ɹ] in young, midwestern, American children," J. Acoust. Soc. Am. **115**, 871–884.

Menn, L. (**1978**). "Phonological units in beginning speech," in *Syllables and Segments*, edited by A. Bell and J. B. Hooper (North-Holland, Amsterdam), pp. 157–172.

Menn, L. (**1983**). "Development of articulatory, phonetic and phonological capabilities," in *Language Production: Development, Writing and Other Language Processes*, edited by B. Butterworth (Academic, New York), pp. 3–50.

Menyuk, P., and Klatt, M. (**1975**). "Voice onset time in consonant cluster production by children and adults," J. Child Lang **2**, 223–231.

Munhall, K. G. (**1985**). "An examination of intra-articulator relative timing," J. Acoust. Soc. Am. **78**, 1548–1553.

Nittrouer, S. (**1991**). "Phase relations of jaw and tongue tip movements in the production of VCV utterances," J. Acoust. Soc. Am. **90**, 1806–1815.

Nittrouer, S. (**1993**). "The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study," J. Speech Hear. Res. **36**, 959–972.

Nittrouer, S., Estee, S., Lowenstein, J. H., and Smith, J. (**2005**). "The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops," J. Acoust. Soc. Am. **117**, 351–364.

Nittrouer, S., Munhall, K., Kelso, J. A., Tuller, B., and Harris, K. S. (**1988**). "Patterns of interarticulator phasing and their relation to linguistic structure," J. Acoust. Soc. Am. **84**, 1653–1661.

Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Hear. Res. **32**, 120–132.

Nittrouer, S., Studdert-Kennedy, M., and Neely, S. T. (**1996**). "How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables," J. Speech Hear. Res. **39**, 379–389.

Oller, D. K., and Eilers, R. E. (**1982**). "Similarity of babbling in Spanish- and English-learning babies," J. Child Lang **9**, 565–577.

Osgood, C. E. (**1953**). *Method and Theory in Experimental Psychology* (Oxford University Press, New York).

Prather, E. M., Hedrick, D. L., and Kern, C. A. (**1975**). "Articulation development in children aged two to four years," J. Speech Hear Disord. **40**, 179–191.

Schwartz, R. G., and Leonard, L. B. (**1982**). "Do children pick and choose: An examination of phonological selection and avoidance in early lexical acquisition," J. Child Lang **9**, 319–336.

Schwartz, R. G., Leonard, L. B., Loeb, D. M., and Swanson, L. A. (**1987**). "Attempted sounds are sometimes not: An expanded view of phonological selection and avoidance." J. Child Lang **14**, 411–418.

Scobbie, J. M., Gibbon, F., Hardcastle, W. J., and Fletcher, P. (**2000**). "Covert contrast as a stage in the acquisition of phonetics and phonology," in *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, edited by Michael Broe and Janet Pierrehumbert (Cambridge University Press, Cambridge), pp. 194–207.

Sjölander, K., and Beskow, J. (**2000**). "Wavesurfer: An open source speech tool," Proceedings of the Sixth International Conference on Spoken Language Processing, Vol. **4**, 464–467.

Smith, A., and Goffman, L. (**1998**). "Stability and patterning of speech movement sequences in children and adults," J. Speech Lang. Hear. Res. **41**, 18–30.

Smith, B. L., and McLean-Muse, A. (**1986**). "Articulatory movement characteristics of labial consonant productions by children and adults," J. Acoust. Soc. Am. **80**, 1321–1328.

Smith, B. L., Sugarman, M. D., and Long, S. H. (**1983**). "Experimental manipulation of speaking rate for studying temporal variability in children's speech," J. Acoust. Soc. Am. **74**, 744–749.

Snow, D. (**1997**). "Children's acquisition of speech timing in English: A comparative study of voice onset time and final syllable vowel lengthening," J. Child Lang **24**, 35–56.

Studdert-Kennedy, M. (**1987**). "The phoneme as a perceptuomotor structure," in *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing*, edited by A. Allport, D. G. MacKay, W. Prinz, and E. Scheerer (Academic, Orlando), pp. 67–84.

Tyler, A. A., and Saxman, J. H. (**1991**). "Initial voicing contrast acquisition in normal and phonologically disordered children," Appl. Psycholinguist. **12**, 453–479.

Velten, H. V. (**1943**). "The growth of phonemic and lexical patterns in infant language," Language **19**, 281–292.

Vihman, M. M., and Velleman, S. L. (**1989**). "Phonological reorganization: A case study," Lang Speech **32**, 149–170.

Vihman, M. M., and Velleman, S. L. (**2000**). "The construction of a first phonology," Phonetica **57**, 255–266.

Werker, J. F., and Tees, R. C. (**1999**). "Influences on infant speech processing: Toward a new synthesis," Annu. Rev. Psychol. **50**, 509–535.

Whalen, D. H., Levitt, A. G., and Goldstein, L. M. (**2007**). "VOT in the babbling of French- and English-learning infants," J. Phonetics **35**, 341–352.

Zlatin, M. A., and Koenigsknecht, R. A. (**1976**). "Development of the voicing contrast: A comparison of voice onset time in stop perception and production," J. Speech Hear. Res. **19**, 93–111.