Journal of Phonetics (1974) 2, 195–210

Early apical stop production: A voice onset time analysis

Diane Kewley-Port

Haskins Laboratories, New Haven, Connecticut, U.S.A.

and

Malcolm S. Preston

Department of Pediatrics, The Johns Hopkins University School of Medicine, The John F. Kennedy Institute, Baltimore, Maryland, U.S.A.

Received 8th February 1974

Abstract: Voice onset time (VOT) has been shown to differentiate effectively the phonemic categories of stop consonants along the voicing dimension. This study involved the measurement of VOT from spectrograms of apical stops produced by young children acquiring American English. Stops were measured from three children recorded regularly between 1 and 2 years of age and from additional children ranging in age from 6 months to 4½ years. Frequency distributions of apical stops along the VOT continuum are compared longitudinally across subjects and with distributions of adult productions of word-initial /d/ and /t/. Drawing on a physiological discussion of the control of timing between stop release and onset of vocal fold oscillation, a pattern of apical stop development is proposed. The earliest examples of stops, around 6 months of age, have uniform distributions along the VOT continuum. Later the distributions of apical stops collapse into an interval corresponding to that of American English /d/. With further development some apical stops are added in the range of adult /t/. The distributions for both /d/ and /t/ words are similar between the ages of 2 and 4½ years, but they do not yet correspond to those of adults.

Introduction

This investigation applies acoustic measurement techniques to a developmental study of speech production. The measure selected for this study of stop consonants is voice onset time (VOT), which is defined as the time interval between the release of stop occlusion and the onset of vocal fold oscillation. VOT can be easily measured from spectrograms of adult consonant-initial vocalizations. VOT measurements roughly comparable to those of adults can be made from spectrograms of the vocalizations of young infants if criteria for the selection of stop consonants are carefully specified.

Linguists have claimed that voicing is a primary phonetic dimension for distinguishing among categories of stops produced at the same point of articulation. The voicing dimension for stops has been related to many different acoustic and articulatory phenomena. Lisker & Abramson (1971, p. 770) have stated that voice onset time is "the single most effective measure" for sorting stops into different phonemic categories with respect to voicing. Comparisons between different categories of stops can be made by displaying the VOT measured from many repetitions of stop consonants as frequency distributions of the percentage of occurrence of stops along a VOT continuum. Lisker & Abramson (1964, 1967, 1970) have repeatedly shown the reliability of VOT alone for differentiating stop categories in different languages. They have also shown the effectiveness of VOT as a perceptual cue for the voicing distinction for speakers of several languages (Abramson & Lisker, 1965, 1970, 1972).

The present study is an investigation of how the voice onset time distribution can be used to indicate changes in the development of stop consonant production for very young children. A detailed case study of the vocalizations from three children recorded at regular intervals between 6 months and 2½ years of age is presented. Because these children produced almost exclusively apical stops during the second year of life, we report here on only the occurrence of the apical stops. In addition we corroborate some of the longitudinal data with apical stops recorded from other children ranging in age from 6 months to 4½ years. A brief study of the clearly identified words from among all apical stops is also included. The data reported are evaluated in the context of a general discussion of the physiology of stop production and a pattern of development of apical stops as reflected in VOT distribution is proposed.

Method

The data for the longitudinal case studies consisted of three sets of tape-recorded sessions, each set corresponding to one of three normally developing children (E3, E4 and E7) from American English-speaking environments. Tape recordings of E3 were analyzed at 45, 51, 60, 73, 81, 97 and 101 weeks of age. For E4 the ages of analysis were 50, 64, 82, 96, 111 and 125 weeks. For E7 sessions were analyzed at 34, 40, 51, 64, 75, 83 and 96 weeks. These ages were chosen to correspond across subjects at roughly 12-week (3-month) intervals. Recordings having the greatest amount of vocalization were chosen when more than one recording was available for a time period.

E3 was a male, while E4 and E7 were females. E3 and E4 were the children of medical residents at the Johns Hopkins Hospital and E7 was the child of a senior undergraduate at the John Hopkins University. Thus, all three came from educated, middle-class families. Except for occasional colds, the three infants were in good health over the period the recordings were made.

The tape-recording sessions were conducted in a sound-isolated booth (IAC model 1203) with the mother or father and occasionally an experimenter present. The instructions to the parents were simply to encourage the child to vocalize as much as possible. Quiet toys and objects of interest were present during the recording sessions, which generally lasted about 30 min each. The children's vocalizations were recorded at 7½ in/s on an Ampex tape recorder (model AG 350). A condenser microphone and cathode follower (Bruel and Kjaer models 4131 and 4133) were connected by cable to the tape recorder outside the booth.

The procedure for analysis involved a transcription by the first author of the entire session using a modified version of the Peterson-Shoup articulatory phonetic theory (Peterson & Shoup, 1966). (Two sessions, which had unusually large numbers of infant vocalizations, were only partially transcribed.) Although phonetic transcriptions of infant vocalizations are obviously necessary to identify the stop consonants appropriate for measurement, the referent of any symbol in that transcription is unclear. Phonetic theories, such as that of Peterson & Shoup (1966), have been developed for the purpose of describing the phone types of the linguistic vocalizations of adults and are based on substantial knowledge of adult acoustics and articulation and of the correspondence between the two. However, far less is known about the articulatory or acoustic properties of the vocalizations of infants, nor is anything known about the reality of the articulatory mechanisms implied by adults ascribing phone types to the vocal sounds of such young children. Hence, at best our phonetic transcriptions must be considered to be a set of adult phone types which seemed most similar to the vocalizations produced by our infant subjects. It is our belief, however, that our phonetic transcriptions are adequate for the purpose of reliably identifying initial stop consonants produced by young children.

Another problem encountered was to select from the children's recordings a set of vocalizations that would be at least roughly comparable to words with initial stop consonants, as spoken by adults. In order to do this, rigorous selection criteria were developed based on the articulatory parameter values of the Peterson–Shoup theory. A vocalization was considered for analysis as long as its initial portion consisted of a stop consonant and a vowel. The transcriber then carefully judged each one as follows.

For the stop, the primary parameters required were plosive, alveolar (apical), and stop. The secondary parameters specified were: pulmonic air mechanism, egressive air direction, nonfrictional airflow, oral airpath (nonnasal), nonlateral lingual airpath, open pharynx shape, natural tongue body shape (nonpalatalized and nonvelarized), and nonretroflexed tongue apex. Because infants exhibit a notable lack of control with reference to several secondary parameters, flexible criteria were used. Air pressure, whether lenis, normal, or fortis, was not judged except where it might have contributed to an excessively frictional airflow. The type of release, as relating to aspirated, unaspirated, or phonoaspirated stops, and lip shape were not judged. Laryngeal action was judged only for the vowel.

For the vowel, any horizontal and vertical place of articulation with pulmonic air mechanism, egressive air direction, and nonfrictional airflow was accepted. Air pressure, general air path, lingual air path, pharynx shape, tongue shape, apex shape, and lip shape were not judged. Vocal fold oscillation presented a special problem for infants. It would have been impossible to choose as normal any particular kind of oscillation, since vocal fry and falsetto voice were frequently produced by all three infants. Thus the laryngeal actions accepted included breathy, voiced, laryngealized, pulsated, and phonoconstricted; however, voiceless, whispered, constricted, and stopped laryngeal actions were not accepted. Further consideration was not given to portions of a vocalization following the stop and vowel.

Following transcription, wide-band spectrograms were made of the selected vocalizations on a sound spectrograph (Voiceprint model 4691A). To facilitate the measurement process, the vocalizations were always analyzed at half speed. Using the spectrograms, stops were categorized as initial in a small number of ambiguous cases by assuring that the stop was preceded by a pause of at least 50 ms. Stops were discarded where the onset of voicing or release was difficult to identify on the spectrogram.

Measurements of VOT to the nearest 10 ms were taken directly from the spectrograms. VOT is measured as the interval between the first vertical striation, representing glottal pulsation, and the onset of energy ("burst"), representing the release of stop occlusion. When the glottal pulses precede the stop release (voicing lead), the VOT value is given a negative sign; when the stop release precedes the glottal pulses (voicing lag), the VOT value is positive. The second author (or occasionally an assistant) checked the VOT measurements and verified the stop consonant transcription by listening to the tape recordings. Thus, only sounds which were clearly identified as apical stops in initial position and which could be measured for VOT were included in the final analysis. The number of apical stops per session included in the final analysis varies from 13 to 98. However, only three of the total 20 sessions had fewer than 20 tokens.

Results

Figure 1 presents the combined data for each of the three subjects in the form of frequency distributions covering the entire period investigated. Distributions for adults borrowed from the work of Lisker & Abramson (1967, p. 13) are also presented in Fig. 1 for comparison. Their two distributions are derived from sentences, some of which contained words starting with the phonemes /d/ or /t/, spoken by ten American English speakers. The data for each child are presented as one apical stop distribution since there was no way to assign phonemic units to their babbling. Each of the distributions from the children should be compared separately with the distributions of /d/ and /t/ for the adults, as well as with each other. It is evident that the children's distributions are remarkably similar to one another. Each has a single mode, and the majority of the productions fall in the 0 to +20 ms voicing lag region.

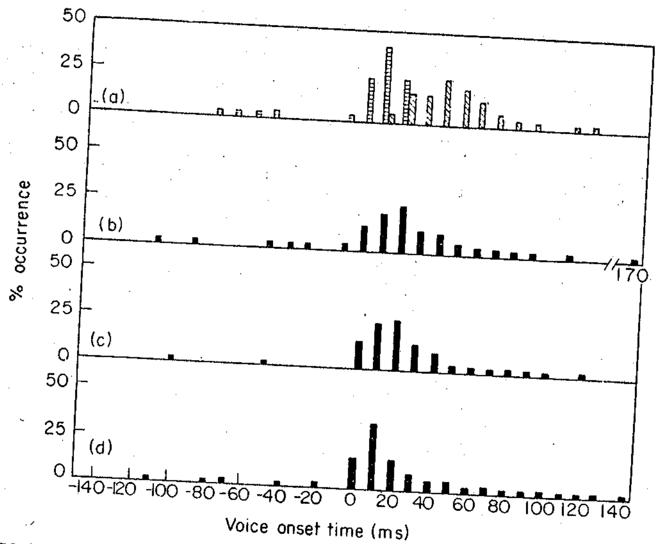

Figure 1 Voice onset time distributions of apical stops for each child combined over the time period investigated and the adult distribution from Lisker & Abramson (1967). (a) L and A, 1967; ▤ /d/, ▧ /t/. (b) E7, 34 to 96 weeks, $N=328$. (c) E4, 50 to 125 weeks, $N=353$. (d) E3, 45 to 101 weeks, $N=266$.

To facilitate comparison of the children's data with those of adults, we introduce some terminology from the studies of Lisker & Abramson (1964, 1970, 1971). In their cross-language studies of initial stop consonants, three categories of stops having a rough correspondence across languages emerge along the VOT continuum. The categories are defined as follows: *voicing lead*, where stops have negative VOT values; *short voicing lag*, where stops have VOT values from 0 to +20 ms; *long voicing lag*, where stops have VOT values greater than +40 ms. As Fig. 1 shows, measurements of American English apical stops produce two partially overlapping frequency distributions with a boundary between +20 and +30 ms. Since the majority of VOT values for /d/ lie in the short voicing lag category and only a small percentage occur in the voicing lead category, it will sometimes be convenient to use the term "d-range" to refer to VOT measurements of +20 ms and less. Similarly, the term "t-range" will refer to VOT values of +30 ms and greater, noting that most values for /t/ lie in the long voicing lag category. The d-range and t-range, as defined, reflect a basic attribute of the voice onset time models which the child will eventually acquire for distinguishing words beginning with /d/ and /t/, namely that values along the voice onset time continuum are divided into two reasonably distinct classes.

A comparison between the children's and the adults' distributions suggests that the children's stops reflect the American English use of both /d/ and /t/. Only about 5% of the apical stops have voicing lead, whereas 64% are in the short voicing lag category and 31% in the long voicing lag category. Thus, children's stops fall in the d-range twice as often as the t-range. The children's distributions have one mode; in contrast, the adult data, if combined into a single distribution, would show two modes, one for each category of apical stop.

Figures 2, 3 and 4 present the data arranged in longitudinal fashion for E3, E4 and E7, respectively. Each distribution in these figures corresponds to a recording session, going from youngest at the bottom to oldest near the top. The Lisker & Abramson distributions for adults are again reproduced at the top of each figure.

Inspection of the data for E3 at 45 and 51 weeks shows a concentration of apical stops in the short voicing lag category, with only a few tokens in the long voicing lag category. By 101 weeks, E3 shows a considerable number of long voicing lag stops ranging from +30 to +160 ms, with no preference for any particular value. There are almost no stops in the voicing lead range at any age. The mode of the distributions falls at +10 or +20 ms VOT.

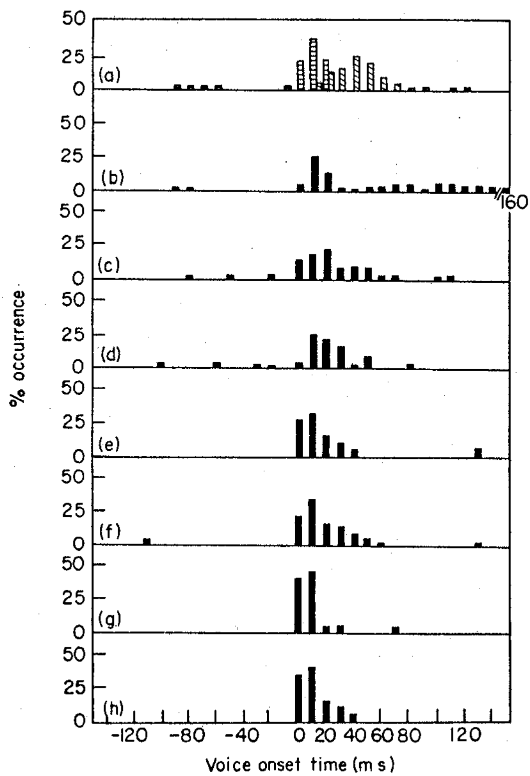

Figure 2 Voice onset time distributions of apical stops for subject E3 for each recording session and the adult distributions from Lisker & Abramson (1967). (a) L and A, 1967; ⊟ /d/, ■ /t/. (b) 101 weeks, $N = 52$. (c) 97 weeks, $N = 39$. (d) 81 weeks, $N = 32$. (e) 73 weeks, $N = 18$. (f) 60 weeks, $N = 45$. (g) 51 weeks, $N = 18$. (h) 45 weeks, $N = 25$.

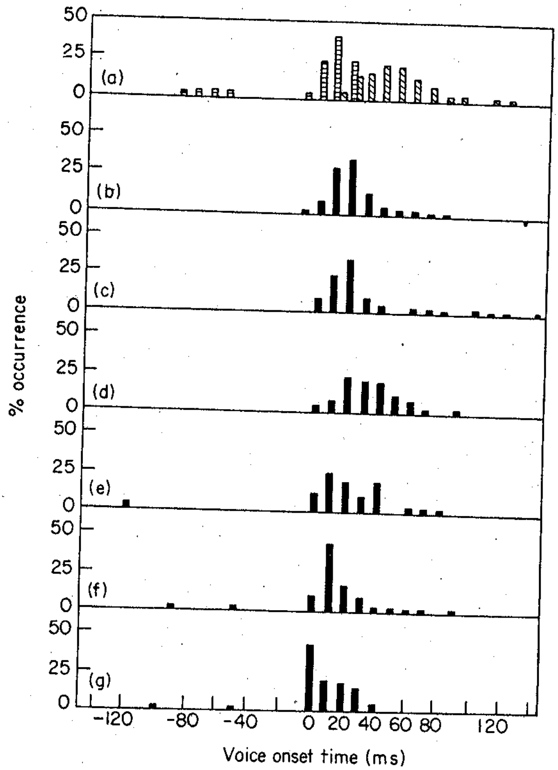

Figure 3 Voice onset time distributions of apical stops for subject E4 for each recording session and the adult distributions from Lisker & Abramson (1967). (a) L and A, 1967; ▤ /d/, ▩ /t/. (b) 125 weeks, $N = 98$. (c) 111 weeks, $N = 71$. (d) 96 weeks, $N = 57$. (e) 82 weeks, $N = 34$. (f) 64 weeks, $N = 36$. (g) 50 weeks, $N = 57$.

For E4 the developmental pattern is much like that of E3. At 50 weeks of age, there is a concentration of short voicing lag stops; although stops do occur in the other categories. By 96 weeks, E4 produces a considerable number of long voicing lag stops and continues to do so at 111 and 125 weeks. Again few stops with voicing lead occur. Distributions have a single mode that ranges from 0 to +20 ms.

The developmental sequence for E7 contrasts in certain respects with that of E3 and E4. First, E7 was recorded at a much earlier age than E3 or E4. The distributions for the earliest recording sessions (34 and 40 weeks) have a wide range of VOT values with no apparent mode. At 51 weeks, unlike E3 and E4, there is still a wide range of VOT values, −90 ms lead to +280 ms lag, although there is now a mode at +20 ms. Thereafter the mode falls at +10 or +20 ms voicing lag. A concentration of stops in the short voicing lag category does occur at 75 weeks, and then at 83 and 96 weeks long voicing lag stops again appear more frequently. E7 has more stops in the voicing lead range than E3 or E4 up to 51 weeks; after 64 weeks such stops also occur infrequently.

These data can be collapsed by categorizing stops into the d-range or t-range as previously defined. Thus a graphic representation of stops in the t-range as a percentage of the total number of stops at each age characterizes the developmental sequence in which stops representative of the adult models of /d/ and /t/ are observed.

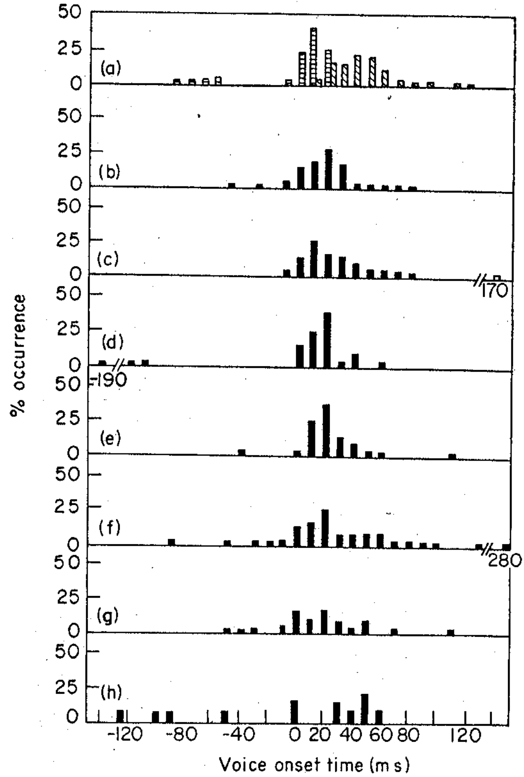

Figure 4 Voice onset time distributions of apical stops for subject E7 for each recording session and the adult distributions from Lisker & Abramson (1967). (a) L and A, 1967; ▤ /d/, ▩ /t/. (b) 96 weeks, $N = 65$. (c) 83 weeks, $N = 55$. (d) 75 weeks, $N = 38$. (e) 64 weeks, $N = 32$. (f) 51 weeks, $N = 94$. (g) 40 weeks, $N = 33$. (h) 34 weeks, $N = 13$.

Graphs of this type for E3, E4, and E7 are presented in Figs. 5, 6 and 7. Subjects E3 and E4 have a similar developmental pattern between 1 and 2 years of age. At about 1 year, only 15% of stops produced are in the adult t-range. This percentage gradually increases until by 2 years the percentage is over 50%. The drop in the percentage by E4 at 111 and 125 weeks—which is represented on the graph by a dashed line—was the result of a distinct change in vocal behavior. Before 2 years of age, vocalizations during the half-hour record-

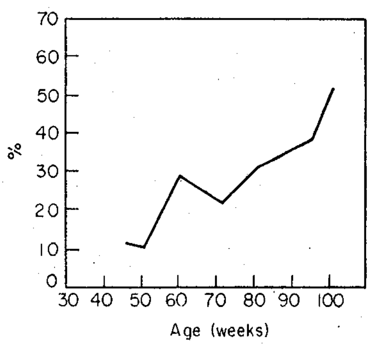

Figure 5 Percent of apical stops with VOT values greater than +25 ms for subject E3.

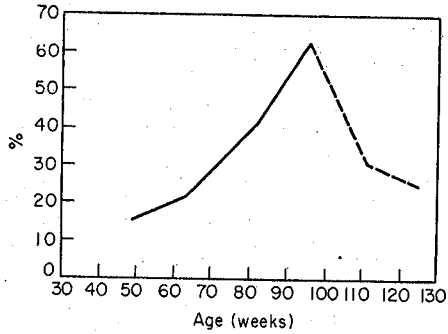

Figure 6 Percent of apical stops with VOT values greater than +25 ms for subject E4.

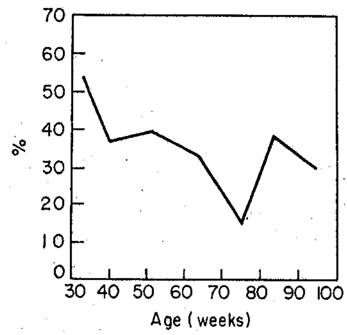

Figure 7 Percent of apical stops with VOT values greater than +25 ms for subject E7.

ing sessions were partially babbling and partially recognizable speech with the attention of the child constantly changing. In later sessions, however, almost all vocalizations were recognizable speech and E4's attention was centered through almost the entire session on a single play activity which happened to involve "dishes". Thus the percentage of stops in the t-range for the older sessions is representative of data that are qualitatively different from that of the younger sessions.

Chronologically, the pattern of development for E7 is not similar to those of E3 and E4. The broad distribution of VOT values observed at 34 weeks is divided half into the t-range, half into the d-range. The percentage in the t-range then falls to 12% at 75 weeks. The percentage increases in following sessions, but at almost 2 years is only 30% compared to over 50% for E3 and E4.

Although there are chronological differences between E7 and the other children, we can interpret the data for all three children from a developmental point of view. From observations of E7 until she was 2½ years old—a time period extending that of data collection—the experimenters observed that the overall language development of E7 lagged considerably behind that of E3 and E4. Taking this into account in the discussion, we propose a pattern of development of apical stops which describes the longitudinal data obtained from all three children.

Discussion

It seems appropriate to relate these case studies to the literature concerned with the physiology of stop consonant production for adults. We will rely extensively on Rothenberg's (1968) carefully developed model of stop consonant production. Further, a large body of more recent data using electromygraphic and fiberoptic techniques will be fit into his basic model. Our particular interest, however, is to suggest a new hypothesis which will be helpful in interpreting the infant data. The hypothesis is that short voicing lag stops are in specific ways easier for the infant to produce successfully than the other two types.

At least three separate articulatory gestures with separate innervations are needed to produce a stop consonant; these are the articulations to permit stop closure and release, to isolate the nasal cavities at the velum, and to initiate vocal fold oscillation. Other articulatory gestures in the vocal tract may also be used by adults to produce stops. However, from the point of view of an infant learning to produce stops, it would appear that control at the point of articulation, the velum, and the larynx must necessarily come first.

The authors agree with the position of Lisker & Abramson (1964, 1967, 1971) and Rothenberg (1968) that the contrastive differences in the voicing dimension of stops are primarily the result of differences in the timing of glottal articulation relative to supraglottal articulation. We propose that distinct physiological mechanisms underlie the production of stops within each of the three voice onset time categories and, further, that stops in the short voicing lag category are easier to produce than stops in the other two categories.

First we will examine the hypothesis that the infant needs to learn only one type of apical articulatory gesture for the production of apical stops, regardless of the VOT category. Although there is rather little data for adults, these studies do not reveal any essential differences in effecting articulatory closure for initial stops differing with respect to voicing. In a palatography study by Fujii (1970) of the dynamic placement of the tongue against the palate, Japanese /d/ and /t/ were considered to belong to a single articulatory class (compared to other consonants), although there were small, consistent differences between them. In other studies, all of labial stops, Harris, Lysaught & Schvey (1965) and Fromkin (1966) investigated electromyographic (EMG) signals from the primary muscle of articulation for labials, the orbicularis oris, and found only insignificant differences in peak EMG strength for English /p/ and /b/. Lubker & Parris (1970, p, 632), using simultaneous measurements of EMG and force of labial contact, found the labial gestures for /p/ and /b/ "essentially monotypic". Measurements of closure duration for American English /p/ and /b/ by Lubker & Parris (1970), and for Dutch /p/ and /b/ by Slis (1970), found durations to be the same in initial position, varying from 100 to 150 ms depending on context. Thus, as inconclusive as these data are, we will assume that an infant could learn essentially one type of apical articulation and be able to produce stops in all VOT categories.

The nasal cavities must be isolated from the rest of the vocal tract in order to create the intraoral pressure needed to produce a stop. Muscles attached to the velum and pharyngeal muscles act to close the velum against the pharyngeal wall. Many recent investigations have shown some differential activity in the velopharynx for stops belonging to different VOT categories (Berti & Hirose, 1972; Lubker, Fritzell & Lindqvist, 1970). The relevance of these studies will be discussed below.

For stop consonants in initial position, the glottal articulation that must be effected is the adduction of the vocal folds from an open (rest) position to a closed, oscillatory position. Voice onset time measurements reflect the time at which the adduction of the vocal folds is achieved relative to the stop release. For apical stops in the short voicing lag category, VOT measurements range from 0 to +20 ms. Direct observation of the larynx by fiberoptic techniques (Lisker, Sawashima & Abramson, 1970; Sawashima, Abramson, Cooper & Lisker, 1970), confirm that the vocal folds have fully adducted and are oscillating at or very near the time of stop release. Thus, articulatory gestures required to produce short voicing lag stops are velopharyngeal closure followed by the complete adduction of the vocal folds at the time of release of the supraglottal articulators, such that vocal fold oscillation begins within 20 ms of release.

In order to initiate vocal fold oscillation, another factor must be coincident. Oscillation of adducted vocal folds is the result of airflow through the glottis which in turn occurs when there is a sustained pressure drop across the glottis. When the vocal tract is unobstructed and the vocal folds are adducted, a wide range of transglottal pressure differentials and tensions in the vocal folds will result in some sort of vocal fold oscillation. However, when the vocal tract is obstructed, as during stop closure, and the vocal folds adducted, Rothenberg (1968, p. 91) has argued that oscillation will not occur or be maintained unless special articulatory mechanism are employed to sustain a transglottal pressure drop. These mechanisms may include active or passive enlargement of the supraglottal cavity, some nasal airflow, and heightened subglottal pressure. Thus, if the vocal folds are adducted at any time during apical closure and additional muscle gestures are not made, vocal fold oscillation will not begin until after the stop closure is released.

Thus, for an infant to successfully produce short lag apical stops in initial position, he may fully close the glottis any time during apical closure providing that the velopharyngeal closure merely isolates the nasal cavities. However, to produce voicing lead stops, the infant must complete glottal closure considerably before oral release and then initiate and sustain vocal fold oscillation by the addition of other articulatory mechanisms (suggested above). These might include velopharyngeal adjustments other than simple velopharyngeal closure.

Stops with long voicing lag are produced with the glottis open at the time of release according to fiberoptic investigations (Lisker, Sawashima & Abramson, 1970; Sawashima, Abramson, Cooper & Lisker, 1970). For American English /t/, the onset of vocal oscillations in the Lisker & Abramson (1967) data in Fig. 1 has a mean of +45 ms VOT. Lisker, Sawashima & Abramson (1970) show that the vocal folds become fully adducted a short time (approximately 30 ms) after oscillation has begun. Kim (1970, p. 111) and other researchers indicate it takes about 100 to 120 ms to adduct the vocal folds fully from their initial open position. These data suggest that an infant will successfully produce a long voicing lag stop if he leaves the glottis open throughout apical closure and then initiates vocal fold adduction approximately at stop release, having maintained velopharyngeal closure throughout. We note that the gesture for velopharyngeal closure could be approximately the same for the infant to produce short and long voicing lag stops, but it is likely to be different and more complex for the voicing lead stops.

The range along the VOT continuum covered by the different VOT categories is also of interest. For the long voicing lag stops, English /t/ can serve as a representative example; Fig. 1 shows 90% of the /t/ stops falling within a 50 ms VOT interval. This indicates that the adult articulation of /t/ involves the very careful control of timing between the supraglottal and glottal articulators, which, we repeat, are separately innervated.

This precise timing constraint is not necessary for the short lag stops. Although the VOT range is about 20 ms, adduction of the vocal folds can be achieved any time during apical closure—approximately 100 ms—and oscillation will still begin only upon apical release. For those languages containing voicing lead stops in the Lisker & Abramson (1964) study, the range of VOT values was approximately 90 ms. Thus, timing between glottal and supraglottal articulators seems more carefully controlled for long voicing lag stops than for voicing lead or short voicing lag stops.

Further support for our hypothesis comes from EMG studies by Hirose & Gay (1972). They studied the intrinsic muscles of the larynx and presented the results for two utterances with contrastive initial stops, /pʌpə/ and /bʌpə/. The adduction of the vocal folds by the interarytenoid muscle—considered to be the primary adductor—for /b/ occurred over a longer period of time and with less force than for /p/. Under different conditions of data collection, Hiroto, Hirano, Toyozumi & Shin (1967, p. 871) came to a similar conclusion: the time interval from the beginning of change of EMG activity to onset of voicing for the interarytenoids was greater for a set of Japanese short voicing lag consonants (including /b/ and /d/) than for a corresponding set of long voicing lag consonants (including /p/ and /t/).

Examining a different muscle, Hirose & Gay (1972) showed that there was substantial motor unit activity for the posterior cricoarytenoid muscle before /p/ but little before /b/ in the pair /pʌpə/ and /bʌpə/. Thus, these EMG studies indicate that muscular activity to adduct the vocal folds is more complex for long voicing lap stops than for short voicing lag stops.

In summary, beginning with the articulatory models of stop consonant production as developed by Rothenberg (1968) and supplemented by the research outlined above, we believe that the articulatory gestures underlying short voicing lag stops are in specific ways less complicated than for the other types of stop. Voicing lead stops require muscle gestures in addition to those needed for short voicing lag stops. The long voicing lag stop requires more carefully controlled timing between stop and laryngeal closures and the adduction of the vocal folds requires more complex muscular activity. These conclusions, we believe, support our hypothesis that the short voicing lag stops have less complex articulations for the infant to control successfully than have the other two types.

Returning to the results, we emphasize that each child studied showed at some age a strong concentration of apicals in the d-range. For E3 it was 45 and 50 weeks, for E4, 50 weeks, and for E7, 75 weeks. Since there was a large age difference between E7 and the others we checked the reliability of her data at 75 weeks by fully analyzing another recording of E7 at 74 weeks of age. Of the 46 apical stops in the final distribution, 98% were in the d-range. Perhaps the later age of stop concentration in the d-range observed for E7 reflects her overall slower language development previously mentioned.

A possible explanation for this concentration of stops in the d-range is that the children were physiologically incapable of producing any other type. However, E7's data would seem to rule this out. She clearly produced apicals in all three categories before one year of age. Her early distribution could, in fact, be characterized as uniform, that is, having apicals randomly distributed over a VOT range from voicing lead to long voicing lag.

Since comparable chronological data were not collected from E3 and E4, we recorded another child, E10, at weekly intervals from 26 to 31 weeks of age. Unfortunately, this child produced almost exclusively velar stops during this period, and usually not very many per recording. At 30 weeks, however, 23 velar stops were analyzed, producing a distribution that was distinctly uniform from −160 ms voicing lead to +160 ms voicing lag.

Thus, we tentatively suggest that children do produce stops in all the VOT categories at some young age, but that at some later time they produce stops only in the short voicing lag category. But we can trace the development of stop production further back in time. It is known that new-born infants do not produce any stops in their vocalizations (Lieberman, Harris, Wolff & Russell, 1971). Our own observations of E7 and E10 found that not until about 6 months of age did they produce enough stops to record even a small number in a half-hour session. It appears then, that the broad, uniform VOT distributions obtained from E7 and E10 characterize infants' initial attempts to produce stop consonants.

We present, then, the following interpretation of the data above. When the infant first produces apical stop consonants, he demonstrates so little control over the timing between the articulatory gestures at the point of articulation and in the glottis that a very broad range of VOT values having no mode is observed. The child will, of course, eventually acquire control over his articulators so that he correctly imitates the VOT models of American English /d/ and /t/. However, when the child first evidences some control only the short voicing lag apical is observed. We suggest that articulatory control for short voicing lag apicals is achieved first because they are less complex articulations for the infant to produce successfully than are the long voicing lag apicals.

Following the period of time when a concentration of short voicing lag apicals was observed, the results showed that long voicing lag apicals appeared more and more frequently. Few voicing lead apicals were produced. Apicals having VOT values in the t-range did not, however, have a mode characteristic of adult /t/. To compare further the adult distributions to the children's, a small study of the recognizable /d/ and /t/ words was carried out.

E4 was the child recorded the longest so her data were examined for words beginning with /d/ or /t/. Words were accepted if two adults could easily identify them as English words beginning with "acceptable" infant productions of /d/ or /t/, i.e. not obviously in error with respect to voicing. The entire utterance in which the /d/ or /t/ word occurred was written down in its approximate English equivalent (for example, *Doggie see fish*) for presentation to other subjects.

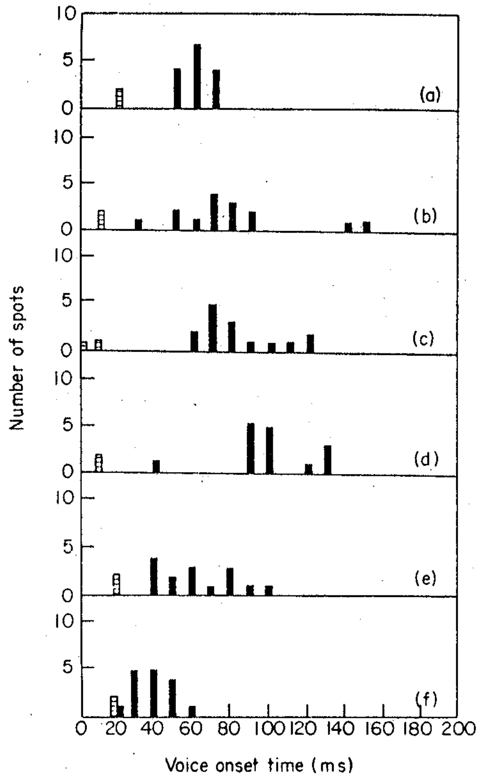

Figure 8 Voice onset time distributions of apical stops for each subject corresponding to the repetitions of the /d/ and /t/ words produced by E4 at 96 weeks of age. ▤ /d/, ■ /t/. (a) JD adult. (b) KI, 4½ years. (c) KP, 4½ years. (d) KE, 4½ years. (e) KM, 3½ years. (f) E4, 96 weeks.

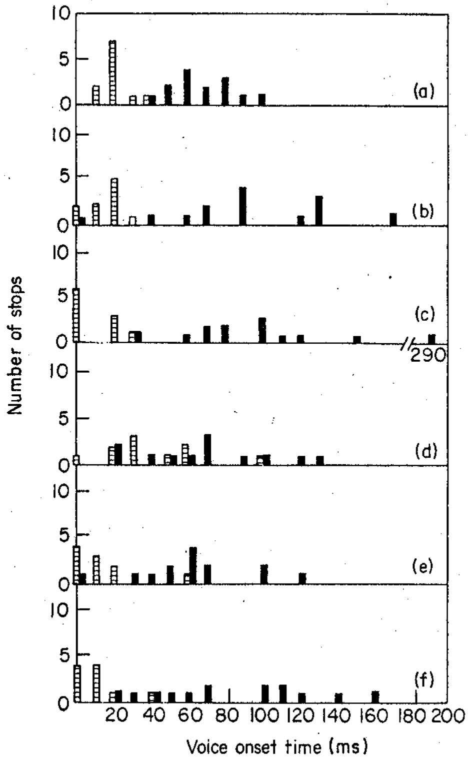

Figure 9 Voice onset time distributions of apical stops for each subject corresponding to the repetitions of the /d/ and /t/ words produced by E4 at 111 weeks of age. ⊟ /d/, ■ /t/. (a) JD, adult. (b) KI, 4½ years. (c) KP, 4½ years. (d) KE 4½ years. (e) KM, 3½ years. (f) E4, 111 weeks.

The earliest age for which we could identify a few /d/ and /t/ words in the half-hour recordings was 96 weeks. We also selected words from recordings at 111 weeks (2¼ years) and 125 weeks (2½ years). For comparison, all of E4's utterances were repeated by four other children and three male adults. The children's ages were 3½ years (one child) and 4½ years (three children). For the children, an experimenter or the mother read from the respective lists of E4's utterances which the child was asked to repeat correctly. The adults simply read the utterance lists. Spectrograms were made of all utterances, and the VOT distributions were made separately for the /d/ and /t/ words for each individual subject. This procedure was used in hope that effects of context would be controlled across subjects. Following analysis, the three adult distributions were essentially identical, so only one is reported here.

The data for all subjects are presented in Figures 8, 9 and 10, corresponding to the utterances for E4 at 96, 111 and 125 weeks of age, respectively. Consider the distributions for /d/ words. Since there were only two /d/ words at 96 weeks, results are mainly based on distributions at 111 and 125 weeks. The /d/ distributions for all five children for both 111 and 125 weeks are the same; thus, remarks on the /d/ data refer in common to Figures 8, 9 and 10. All children's distributions bear basic similarities to those of the adult, with small differences. Adult JD has a VOT range for /d/ of +10 to +40 ms, with only 10% of the /d/s falling in the t-range. E4 has a VOT range of 0 to +40 ms, with 25% of the /d/s falling in the t-range. The four older children have a range for /d/ of 0 to +100 ms, again with 25% of the

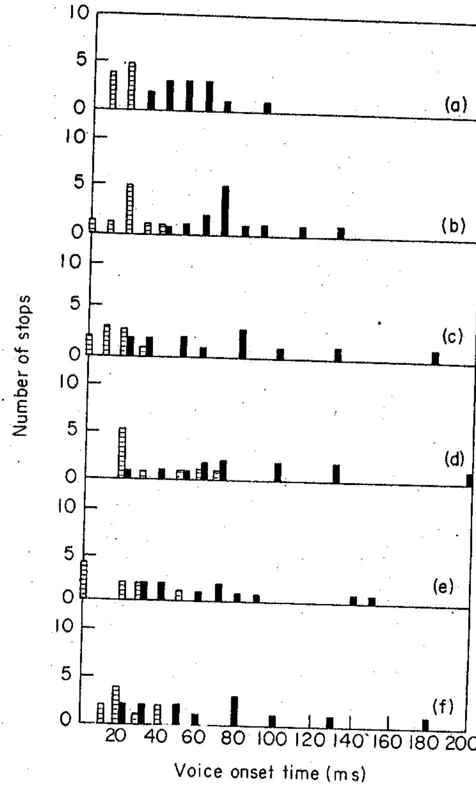

Figure 10 Voice onset time distributions of apical stops for each subject corresponding to the repetitions of the /d/ and /t/ words produced by E4 at 125 weeks of age. ⊟ /d/, ■ /t/. (a) JD, adult. (b) KI, 4½ years. (c) KP, 4½ years. (d) KE, 4½ years. (e) KM, 3½ years. (f) E4, 125 weeks.

/d/s in the t-range. (No /d/s with voicing lead occurred.) We thus conclude that distributions for /d/ are the same from the earliest word productions to at least 4½ years of age. The /d/ distributions are quite similar to that of the adult model, but children show considerably more error in producing /d/ words with VOT values in the t-range.

For the /t/ words, there is a difference in the distribution for E4 at 96 weeks compared to the two older distributions. E4's /t/ words at 96 weeks have a mean of +40 ms and a range of +20 to +60 ms. This mean is significantly smaller than the mean (+60 ms) of the /t/ distribution for JD ($p < 0{\cdot}005$, two-tailed test). Few of E4's /t/ words have VOT values in the d-range. The /t/ distribution at 96 weeks, in fact, lies midway between adult /d/ and /t/ distributions observed in Fig. 1. Therefore, according to listening tests for adults (Abramson & Lisker, 1970), if these stops (for /t/ words) were judged on the basis of VOT alone, they would be ambiguously categorized as /d/ or /t/.

E4's /t/ distribution changes radically by 111 weeks of age and is by then similar to all other /t/ distributions for all the children (except E4 at 96 weeks). The characteristic /t/ distribution has a wide range (0 to +290 ms for all children), but has a mean not significantly different from adult JD's mean of +65 ms for all /t/ distributions. The distributions have no apparent mode, with the possible exception of subject KI (4½ years). For all children, very few of the /t/ words intrude into the d-range. The children's /t/ distributions contrast with that of adult JD which has a narrower range of VOT values, +30 to

+100 ms, and a mode at +60 ms. It is also of interest to note that E4's /t/ words actully account for the major portion of all apical stops collected at 96, 111 and 125 weeks that have VOT greater than +30 ms.

A number of conclusions can be drawn from these results. When children begin to use /t/ words, they distinguish their productions functionally from /d/ words along the dimension of voice onset time. However, VOT distributions for /t/ clearly deviate from the adult model. At the earliest age for which we could collect some /t/ words, E4's distribution lies ambiguously across the boundary between the adult /d/ and /t/ distributions, but with VOT values nonetheless clearly larger than those for the majority of adults' or children's /d/ words. We cannot offer an explanation as to why this distribution is so different from the other children's /t/ distributions. It does not appear to be an artifact. The word *toy* (or *toys*) occurs both at 96 and 111 weeks. For four utterances of *toy*(*s*) at 96 weeks, the range of VOT is +20 to +40 ms; for ten occurrences of *toy*(*s*) at 111 weeks, the range is +20 to +140 ms. Thus, at 96 weeks of age E4 has learned one aspect of the adult model for /t/, that /t/ should be produced with VOT greater than +25 ms. But she has not learned to produce apical stops with a large enough delay in the onset of voicing that they would be unambiguously categorized by adults as /t/ on the basis of VOT alone. However, since our adult listeners judged that the /t/ words clearly began with [t], it is possible that some other cue besides voice onset time was being effectively signaled at this age.

By 111 weeks, however, a new pattern of /t/ production occurs which continues to be characteristic of E4's and of our other children's speech through 4½ years of age. This /t/ distribution characteristically has a wide range of VOT values and no mode. At this stage, two of the most important aspects of the adult VOT model have been acquired: the VOT values are greater than +25 ms, and the vast majority of the VOT values are large enough to be unambiguously categorized as /t/ by adult listeners. However, the adult /t/ VOT range and mode are still to be acquired. We have proposed an explanation for a /t/ range that is wide: the production of /t/ is a difficult and complex articulation in which the timing evidenced by the adult model must be finely controlled. What is surprising, however, is that at 4½ years, two years after E4's 111-week distribution, no particular change in the /t/ distribution occurs. That is, we know that children will eventually control their /t/ productions within the limits of the adult model, but this control has not been achieved as late as 4½ years. Sachs[1] has some roughly comparable VOT data for /p/ which show that by 5 years of age, half of her six subjects have a mode and a more restricted range of VOT values. It appears, then, that acquisition of the adult model for /t/ does not occur until after 5 years of age.

Summarizing, we have traced in some detail the development of apical stop consonants with respect to VOT for three children from 1 to 2 years of age and have reported data from other children whose ages span 6 months to 4½ years. A sequential pattern of the development of apical stops with respect to VOT has been suggested. No stops are observed in neonatal vocalizations. When stops first appear around 6 months of age, the VOT distribution has a wide range of randomly distributed values extending from voicing lead to long voicing lag. Some months later a concentration of apicals in the short voicing lag category is observed. Apical stops in the long voicing lag category are then gradually added. When words beginning with /d/ and /t/ are first observed—about 2 years of age for subject E4—the characteristics of these distributions remain constant until at least 4½ years. The distribution for /d/ looks similar to the adult's distribution but with more errors into the t-range. The /t/ word distributions bear less resemblance to the adult's, having a wide range of VOT values with no mode, although few errors occur into the d-range.

[1] J. Sachs, personal communication, University of Connecticut (1971).

This study was begun at the Neurocommunications Laboratory, The Johns Hopkins School of Medicine, where it was supported in part through the National Institute of Child Health and Human Development, Contract No. PH 43-65-637. Completion of the study was supported in part by the National Institute of Child Health and Human Development, Grant No. 5 P01 HD01994, and by the Public Health Service, Health Services and Mental Health Administration, Maternal and Child Health Service, Grant No. 917.

The authors would like to thank Arthur S. Abramson and Leigh Lisker for their advice during the research itself and comments on an earlier draft of this manuscript. We are most grateful to Hajime Hirose, Rachel E. Stark, Grace H. Yeni-Komshian, and Jacqueline Sachs for many fruitful discussions and ideas.

References

Abramson, A. S. & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In (D. E. Commins, Ed.) *Proceedings of the Fifth International Congress on Acoustics*, Liege: Imp. G. Thone, **A51**.

Abramson, A. S. & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague*, 1967, Prague: Academia, pp. 569–573.

Abramson, A. S. & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* **1**, 1–8.

Berti, F. B. & Hirose, H. (1972). Velopharyngeal function in oral/nasal articulation and voicing gestures. *Haskins Laboratories Status Report on Speech Research* **SR-28**, 143–156.

Fromkin, Victoria (1966). Neuromuscular specification of linguistic units. *Language and Speech* **9**, 170–199.

Fujii, I. (1970). Phoneme identification with dynamic palatography. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo) **4**, 67–73.

Harris, K. S., Lysaught, G. F. & Schvey, M. M. (1965). Some aspects of the production of oral and nasal labial stops. *Language and Speech* **8**, 135–147.

Hirose, H. & Gay, T. (1972). The activity of the intrinsic laryngeal muscles in voicing control. *Phonetica* **25**, 140–164.

Hiroto, I., Hirano, M,, Toyozumi, Y. & Shin, T. (1967). Electromyographic investigation of the intrinsic laryngeal muscles related to speech sounds. *Annals of Otology, Rhinology and Laryngology* **76**, 861–872.

Kim, C.-W. (1970). A theory of aspiration. *Phonetica* **21**, 107–116.

Lieberman, P., Harris, K. S., Wolff, P. & Russell, L. H. (1971). Newborn infant cry and nonhuman primate vocalizations. *Journal of Speech and Hearing Research* **14**, 719–727.

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

Lisker, L. & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech* **10**, 1–28.

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague*, 1967, Prague: Academia, pp. 563–567.

Lisker, L. & Abramson, A. S. (1971). Distinctive features and laryngeal control. *Language* **47**, 767–785.

Lisker, L., Sawashima, M. & Abramson, A. S. (1970). Cinegraphic observations of the larynx during voiced and voiceless stops. *Haskins Laboratories Status Report on Speech Research* **SR-21/22**, 201–210.

Lubker, J. F., Fritzell, B. F. & Lindqvist, J. (1970). Velopharyngeal function: An electromyographic study. *Speech Transmission Laboratory, Quarterly Progress and Status Report* (Royal Institute of Technology/KTH/, Stockholm) **4**, 9–20.

Lubker, J. F. & Parris, P. J. (1970). Simultaneous measurements of interoral pressure, force of labial contact, and labial electromyographic activity during production of the stop cognates /p/ and /b/. *Journal of the Acoustical Society of America* **47**, 625–633.

Peterson, G. E. & Shoup, J. E. (1966). A physiological theory of phonetics. *Journal of Speech and Hearing Research* **9**, 5–67.

Rothenberg, M. (1968). The breath-stream dynamics of simple-released-plosive production. *Bibliotheca Phonetica* **6**.

Sawashima, M., Abramson, A. S., Cooper, F. S. & Lisker, L. (1970). Observing laryngeal adjustments during running speech by use of a fiberoptics system. *Phonetica* **22**, 193–201.

Slis, I. H. (1970). Articulatory measurements on voiced, voiceless and nasal consonants. *Phonetica* **21**, 193–210.