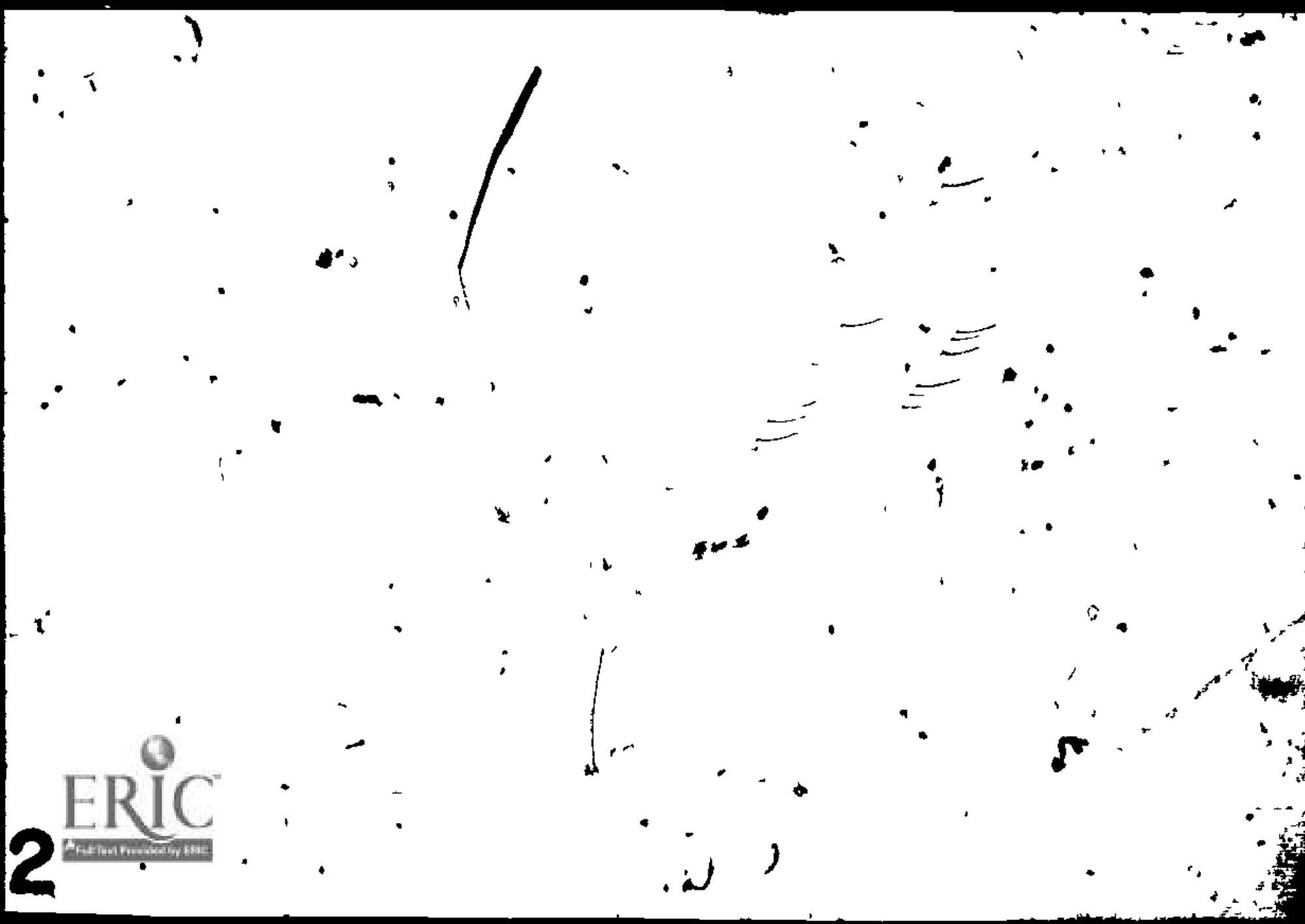
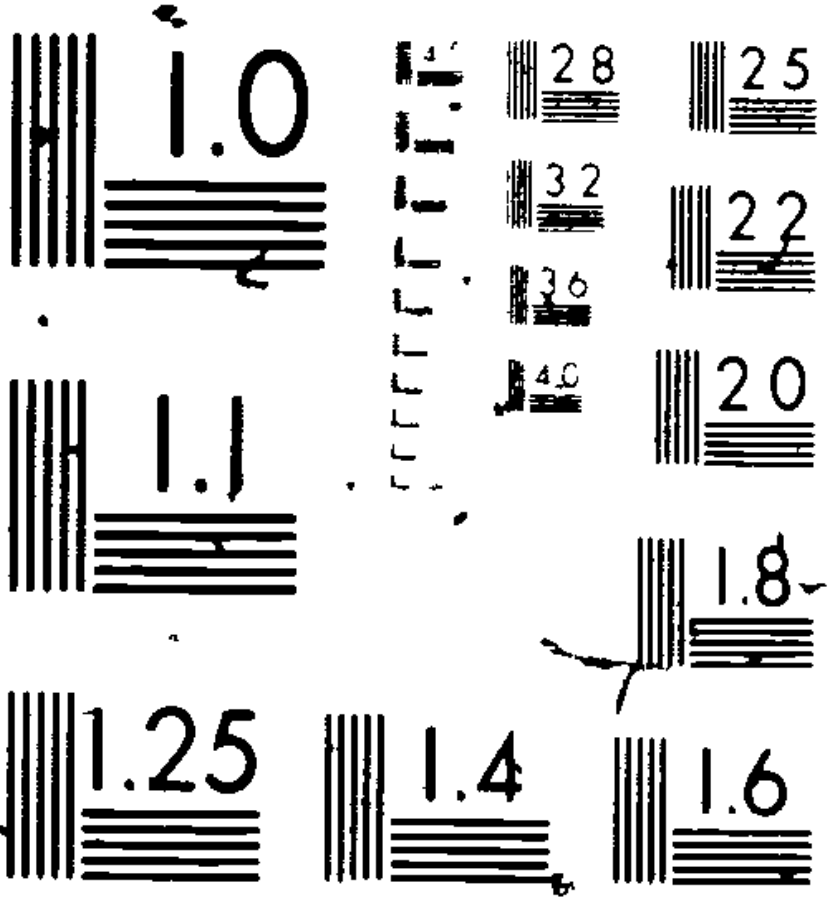




DOCUMENT RESUME

ED 152 084 FL 009 316

AUTHOR Macken, Marlys A.; Barton, David
TITLE The Acquisition of the Voicing Contrast in English: A Study of Voice-Onset Time in Word-Initial Stop Consonants.
PUB DATE Mar 78
NOTE 46p.; Report from the Stanford Child Phonology Project; Best copy available

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Child Language; *Consonants; Descriptive Linguistics; *Distinctive Features; *English; *Language Development; Language Research; Language Skills; Linguistic Theory; Phonemics; Phonological Units; *Phonology; Verbal Development

ABSTRACT

This paper reports on a longitudinal study of the acquisition of the voicing contrast in English word-initial stop consonants, as measured by voice onset time. Four monolingual children were recorded at two week intervals, beginning when the children were about 1;6. Data provided evidence for three general stages: (1) the child has no contrast; (2) the child has a contrast but one that falls within the adult perceptual boundaries of one (usually voiced) phoneme and thus is presumably not perceptible to adults; and (3) the child has a contrast that resembles the adult contrast. The rate and nature of the developmental process are discussed briefly in relation to two competing models for phonological acquisition and two hypotheses regarding the skills being learned. (Author)

Marlys A. Macken and David Barton
Department of Linguistics
Stanford University
3/9/78

The acquisition of the voicing contrast in English: a study of voice-onset time in word-initial stop consonants.

Address correspondence to:

M. A. Macken and D. Barton
Department of Linguistics
Stanford University
Stanford, California 94305
U.S.A.

The acquisition of the voicing contrast in English: a study of voice-onset time in word-initial stop consonants.[1]

Marlys A. Macken and David Barton
Department of Linguistics
Stanford University

ABSTRACT. This paper reports on a longitudinal study of the acquisition of the voicing contrast in English word-initial stop consonants, as measured by voice onset time. Four monolingual children were recorded at two week intervals, beginning when the children were about 1;6. Data provide evidence for three general stages: (1) the child has no contrast; (2) the child has a contrast but one that falls within the adult perceptual boundaries of one (usually voiced) phoneme and thus is presumably not perceptible to adults; and (3) the child has a contrast that resembles the adult contrast. The rate and nature of the developmental process are discussed briefly in relation to two, competing models for phonological acquisition and two hypotheses regarding the skills being learned.

1.0 Introduction

This paper reports on a longitudinal study of the acquisition of the voicing contrast in American-English, as revealed through instrumental analysis of voice onset time characteristics of word-initial stop consonants.

Although the glottal features related to the voicing contrast are among the most controversial in contemporary phonological analysis (Chomsky and Halle 1968, 326-9; Ladefoged 1971, 7-22), it is generally agreed that a single phonological voice distinction in any language may consist of a number of relatively independent phonetic components which nonetheless tend to covary. Among all the articulatory and perceptual elements which are relevant to the phonological voicing contrast, the one which has probably received the most attention is voice onset time (VOT), that interval between the release of stop closure and the onset of vocal fold vibration.

Lisker and Abramson in their seminal 1964 article claim that VOT is the single, most reliable feature separating voiced, voiceless stop cognate pairs in languages of the world, regardless of the conventional feature designations used in descriptions of those languages. Among the eleven languages they studied, three categories of VOT emerged: voicing lead where voicing begins about 75 to 125 milliseconds (ms) prior to burst release; short lag, where voicing begins from 0 to 25 ms after the release; and long lag where voicing begins 60 to 100 ms after the burst. The median values for these ranges are -100 ms, +10 ms and +75 ms respectively (Lisker and Abramson 1964). In English, the voiced phoneme is typically short lag (although some speakers use some amount of voicing lead), while the voiceless phoneme is characterized by long lag.

Because of the Lisker and Abramson claim--which has received substantial support from the findings of other researchers--that VOT reliably distinguishes voicing stop cognate pairs in initial position, and also because there is a large literature documenting the VOT characteristics of voicing in the speech of adult speakers of English, we have chosen VOT as the unit of analysis for our study. We assume then that as the child acquires productive control over voicing, VOT values will change concomitantly over this time period. We will define the presumed goal to which the child is progressing as those characteristics of adult English stop phonemes described in the literature in terms of VOT. The aims of our study are to specify the temporal characteristics of very young children's stop productions, to determine the age at which children acquire the voicing contrast in English and to document the process by which it is acquired.[2]

With respect to the age at which the voicing contrast is acquired, there is wide disagreement in the child phonology literature. The age at which English speaking children acquire voicing in initial stops apparently may vary from 1;3 to 2;8: 1;3 and 1;4 (Moslin 1976); by 1;9 and 1;11 (two subjects in Barton 1976); 2;0 (Leopold 1947); 2;1 (Velten 1943); by 2;2 (one subject in Bond and Wilson 1977); 2;4 (Major 1976); 2;8 (Smith 1973). Dodd 1976 reports that five children no longer made any voicing errors after the age range 2;5 to 2;11. Two recent cross-sectional studies show that many English-

speeking children acquire this contrast by at least 2;6 (Zlatin and Koenigsknecht 1976; and Gilbert 1977). A few studies of children acquiring languages other than English--languages which have at least a contrast between short and long lag stops--report ages for individual children as follows: 1;5, Germen (Lorentz, pers. commun.); 1;7, Garo (Burling 1959); and 2;0, Hindi (Srivistava 1974; specifically voiceless, unaspirated by 1;1, prevoiced by 1;4, voiceless and voiced aspirated beginning at 2;0). Lin 1971 reports that three children acquiring Taiwanese still made errors on the long lag stops at 2;0. The differences between the findings of these studies no doubt in part stem from differences in methodology (e.g. the criterion adopted for determining 'acquisition'), but there are undoubtedly at least some individual differences between the children studied.

In contrast to the disagreement on the age at which the voicing contrast is acquired, there is overwhelming agreement on what type of voicing is used first by children. Jakobson in 1941 accurately observed that "so long as stops in child language are not split according to the behavior of the glottis, they are generally pronounced as voiceless and unaspirated" (1941/1968, 14). In all the studies we surveyed, the children studied first used short lag stops (i.e. 'voiceless and unaspirated' in Jakobson's terminology). Kewley-Port and Preston 1974 claim that short lag stops are used first because they are articulatorily easier to control than either lead or long lag stops.

While the literature substantially demonstrates the priority of short lag stops, different studies provide conflicting data on the question of whether children first acquire a voicing contrast in a particular place of articulation (before other places), and whether the rate of acquisition, once begun, is rapid or slow as the voicing contrast spreads through all places of articulation, all words and all types of speech (e.g. spontaneous versus imitated speech). Towards our goal of documenting the process of acquisition, we examine stops at all three places of articulation, in a variety of different words, and in several types of speech, and we examine developmental changes in the data for evidence supporting the two major, competing models which have been advanced to describe the rate and characteristics of change in phonology acquisition--the 'across-the-board' model (Smith 1973) and the 'lexical diffusion' model (Hsieh 1972 and Ferguson and Farwell 1975). In order to look in depth at the acquisition process, it was necessary to restrict our analysis to word-initial position; thus, we will not provide data on whether voicing is acquired differently in different positions of the word.[3]

2.0 Method

2.1 Subjects

Home interviews with prospective subjects and their families were conducted to determine the degree of cooperation, the verbal level, response to picture books, and approximate vocabulary size and phonological system of the children. Four children were selected to be subjects because they were monolingual speakers of English with no siblings of school age, they were producing at least some initial stop words, they showed evidence of normal language development, and they appeared to be cooperative. Table 1 gives pertinent data on the four subjects, who will here be identified as Tom,

Tessa, Jane and Jay[4]; all subjects are children of professional parents either one or both of whom have advanced degrees.

[INSERT TABLE 1 ABOUT HERE]

Both parents of all the children except Jane are native speakers of American-English. Jane's mother was born in Canada but has spent the last six years in the United States. It was not expected that the slight Canadian accent of Jane's mother (which noticeably only affected some vowels and the stress of some words) would effect Jane's acquisition of voicing. None of the parents speak a language other than English in the home, and none of the children had been exposed to a language other than English.

Tom, Tessa, Jane and Jay were using between 50 and 100 words at the beginning of recording when they were 1;6.24, 1;4.28, 1;6.19 and 1;7.9 respectively (ages are given in years, months and days). Audiometric screening of the children was done by the staff of the Audiology Clinic, Stanford Medical School; all children were assessed as having normal hearing.

2.1 Data collection

The children were seen approximately every two weeks for an eight-month period. All the children were seen weekly for a short period of time at the beginning of the study. One subject, Jay, was seen an extra three times after the study had been officially ended. Since Jay's family moved to Boston at the end of the study, the last session was recorded, under comparable recording conditions, at Boston University.[5] Table 2 gives the age and the number of 'correct' stop-initial words produced by each child on each session. 'Correct' means that the child's production began with a stop of the same place of articulation as the initial stop in the adult word.

[INSERT TABLE 2 ABOUT HERE]

Each session was conducted in a high quality sound-isolated room (with associated control room) at the Stanford Speech Research Laboratory. Recording was made on a Revox A77 tape recorder with a Sony Electret microphone which was attached to a soft cloth vest which the child wore. In most cases, a parent was present. All testing was conducted by the same experimenter, a native speaker of American-English. The observer, who sat in the control room, ran the tape recorder, took notes on the session and kept a tally of the number of stop-initial words produced by the child for all six stop consonants.

Sessions were between twenty and thirty minutes long. The goal for each session was to obtain at least fifteen tokens for each stop consonant. The observer would indicate towards the end of the session that a certain number of words beginning with particular stops were needed. The experimenter would then focus on words beginning with those stops. This number was estimated to be the maximum number of tokens which we could reasonably expect to obtain from children aged 1;6 and the minimum number of tokens required for statistical analysis. During the earliest sessions, it was difficult, however, to obtain ninety target items (fifteen tokens per stop) from three of the four children.

Table 1. Subjects

Subject	Sex	Age		Approximate Vocabulary At Beginning	Siblings	Parents' Dialect Area
		First Session	Last Session			
TOM	M	1;6.24	2;1.3	100 words	Sister (2½ years older)	M: New Jersey F: Utah
TESSA	F	1;4.28	2;0.10	75 words	Brother (2 yrs younger)	M: California F: Philadelphia
JANE	F	1;6.19	2;1.16	75-100 words		M: Canada/California F: California
JAY	M	1;7.9	2;4.2	50 words		M: New York F: New York

Table 2. Age and number of tokens* for each subject by session.

Session	TOM		TESSA		JANE		JAY	
	Age	# Tokens	Age	# Tokens	Age	# Tokens	Age	# Tokens
1	1;6.24	35	1;4.28	106	1;6.19	32	1;7.9	31
2	1;7.2	25	1;5.2	199	1;6.22	56	1;7.16	44
3	1;7.7	66	1;5.18	186	1;6.29	26	1;8.8	46
4	1;7.21	43	1;5.30	130	1;7.12	27	1;8.20	43
5	1;8.29	100	1;6.15	94	1;7.26	33	1;9.4	42
6	1;9.12	53	1;7.6	147	1;8.28	47	1;9.22	53
7	1;9.23	93	1;7.13	99	1;8.24	39	1;10.5	92
8	1;10.9	100	1;7.27	168	1;9.15	77	1;10.15	100
9	1;10.22	117	1;8.10	214	1;10.7	96	1;11.2	83
10	1;11.8	135	1;8.17	142	1;10.14	94	1;11.14	72
11	1;11.22	122	1;9.2	175	1;11.0	140	2;0.0	197
12	2;0.5	118	1;9.23	174	1;11.14	108	2;0.14	176
13	2;0.19	197	1;10.7	217	1;11.28	99	2;0.28	160
14	2;1.3	109	1;10.21	169	2;0.11	56	2;1.12	159
15	2;1.17	139	1;11.11	139	2;0.25	79	2;1.25	162
16			2;0.10	132	2;1.16	94	2;2.2	177
17							2;2.9	135
18							2;3.0	129
19							2;4.2	161

*Tokens = Words which began with an initial singleton stop in the adult model and which were produced by the child with a stop of the same place of articulation.

Stimulus materials included a specially constructed stimulus book, several children's picture books, a small wooden puzzle and many small toys. Word lists were obtained frequently from the parents; objects for or pictures of stop-initial words that each child knew were then added to the stimulus materials. Occasionally the parents brought toys or books with which the child was familiar. During the sessions the children played with the toys, looked at pictures, and were encouraged to speak.

Since we discovered that all the children used few g-initial words (at least at the beginning of the study) and that fewer p-initial than b-initial words were used by at least some of the children, we introduced a few words (e.g. goat, gate, and penny) which were then learned at some point by the children. Particularly for g- and p-initial words, we relied heavily on the parents who would either bring the child's toys (e.g. Piglet) or would help us set up games which would elicit a particular word from an individual child (e.g., driving a car into a pretend garage). The lexical asymmetries for each child will be discussed in the results sections. In general, the majority of words produced by the children were either single syllable words (of the form CVC) or two syllable words with initial syllable stress; in the earliest sessions, all the children produced more b-, d- and k-initial words than words beginning with p-, t- and g- (although the frequency of d- and t-initial words was often similar). The instability of /p/ and /g/ in stop systems of languages of the world has been discussed by Gamkrelidze 1975 and Ferguson 1977.

2.2 Data analysis

All tapes were transcribed by a researcher using a Revox A77 tape recorder and listening under Super St-pro B-V head phones. The tape script included all words that the child said; however, only 'those 'target words' that the child said were phonetically transcribed. Target words were those that began with an initial singleton stop in the adult model and for which the child produced a stop of the correct place of articulation (as judged by the transcriber)[6]; see Table 2 for the total number of such words produced on each session by each child. For these words, the initial phone was transcribed narrowly and the rest of the utterance somewhat more broadly. The transcription system used is that of the International Phonetic Association with a supplemental symbology and diacritic system devised by the Stanford Child Phonetics Workshop (Bush et al. 1973).

Preceding context and the gloss were noted for all productions. The following system of classification was used for all tokens: S = spontaneous; I = imitated; R = repeated (i.e. a repetition of the child's own previous utterance); E = elicited (i.e. the child's response to a question, not containing the target word, asked by the experimenter); and M = modelled (i.e. the child's production of a word spoken by the experimenter after some intervening speech either child- or adult-produced).

Of the biweekly recorded sessions, approximately four anchor point sessions from each subject were selected for instrumental analysis. The purpose behind selecting anchor points was to sample throughout the data. Thus, these anchor points were at approximately two month intervals,

beginning with the first sessions and ending with the last ones. However, in the earliest sessions, not all the children produced the requisite fifteen tokens of each stop consonant; for these children, the first two or three sessions were combined into one anchor point if the sessions had been conducted at close time intervals. Thus, in some cases, data from several sessions were collapsed; these cases are clearly identified in the tables. When two or more sessions were combined, equal proportions of stops were taken from each session.

Wide-band (and in some cases, narrow-band) spectrograms were made of the first fifteen tokens of each stop type on a Kay Sonograph (model 7029 ADC, with an added custom shaping circuit (HS2, 12 db high frequency pre-emphasis)). To achieve better temporal resolution, utterances were recorded on the Kay in the 160-16K mode (which resulted in a time base of 41 ms per centimeter and a scanning filter width of 600 Hz) and were reproduced with scale magnifier set at 0 to 50%. Measurements of VOT were made at the same time on a Tektronix Oscilloscope with storage capability (Type 564), which was used in conjunction with a locally designed and constructed triggering system. Directly following the oscilloscope measurement, VOT was measured on the spectrogram(s). When the oscilloscope and spectrogram measurements differed by more than 3 ms, the problem was identified and then resolved by further instrumental analysis. In all cases where possible the time scale used was 2 ms or 5 ms per centimeter. In general, if the difference between the two measurements was equal to or less than 3 ms, they were averaged to give a final VOT value. However, for VOT values greater than +50 ms (i.e. cases in which the time scale had to be set at 10, 20 or 50 ms per division), fine time resolution was more difficult on the oscilloscope (due to greater compression of the signal in each time division), and greater reliance was placed on the spectrogram measurement. The procedures for instrumental analysis and VOT measurement are described in detail in the report of our pilot study (see Huntington et al. 197").

If an utterance presented a problem which the researcher could not solve, all spectrograms for the item were presented at the weekly group meetings where Project staff (four to six persons) discussed and resolved the problem. In the few cases where the staff could not reach unanimous agreement, the utterance was not included in the statistical analysis. The criteria for rejection of an utterance from VOT measurement were: 'noisy' (e.g., clanging toys during the child's production); 'voice overlay' (where an adult's voice was superimposed over the child's and each could not reliably be separated by narrow-band analysis); 'no burst'; 'following voiceless vowel'; and 'continuous voicing' (where voicing continued uninterrupted from the child's previously voiced segment). Less than 7% of the total set of utterances from all children were rejected.

Each author carried out the transcription and instrumental analysis for two subjects. At approximately two month intervals, reliability checks were made. For these checks, six items were selected (usually the first token of each stop consonant on a tape). Each author transcribed, instrumentally analyzed and measured each item. In several cases, the six items had been analyzed by one or the other author several months previously. Thus, the comparison checks indicated both inter- and intra- observer reliability. In the reliability check procedure (as in the regular analysis procedure), provision was made to allow an observer to label any utterance as a problem which should then be

referred to the weekly group meetings. Three of the 35 reliability check items were identified as problems by both authors independently. Excluding these three items, the two authors agreed to within 6 ms on 97% of the items checked. During the weekly group meetings, Project staff resolved the problems for two of the three problem utterances; the third item was rejected.

For each child, each anchor point was analyzed separately. When definitive VOT values were obtained for fifteen tokens of each of the six stop consonants, separate frequency distributions were drawn and the mean, standard deviation and range calculated. Tests of significance were made of the differences between the mean VOT of the voiced and voiceless phonemes at each place of articulation. Throughout, the probabilities given are for one-tailed t-tests. Where F values are significant, separate variance estimates are used; where they are not significant, pooled variance estimates are used. In general, when the differences between means were significant, they were highly significant, and when they were not significant, they were not at all close to significance. Since the distributions were not completely normal, in the few cases where differences between the means were just significant, a non-parametric test (the Mann-Whitney U test) was also used: in all cases, this test gave the same level of significance as the parametric t-test.

If the differences between the means for a voiced-voiceless pair was significant at one anchor point but not at the preceding point, the intervening sessions were instrumentally analyzed to determine more precisely the point of change. If the differences between the means for a pair were not significant on contiguous sessions or anchor points, the data from these sessions were combined for additional statistical analyses: the difference between the means for each pair from the collapsed sessions were compared (one-tailed t-tests), and a two-way analysis of variance was carried out. Thus, there are two separate cases in which data from separate sessions were collapsed: (1) cases in which the child on early sessions did not produce the requisite fifteen tokens per stop; and (2) cases in which the data did not show a consistent significant difference across several contiguous sessions. The decision to collapse data in the later cases was motivated by the assumption that if no significant change occurred over the time period, the data from that time period represented a single stage and thus were comparable.

3.0 Results

Data from individual children will be discussed in sections 3.1 (Tom), 3.2 (Tessa), 3.3 (Jane) and 3.4 (Jay). Tables 3, 4, 5, and 6 contain the mean VOT, number of tokens, standard deviation and range of productions for each stop consonant on each session analyzed for each of the children. Also included in these tables are the significance levels for the difference between the means for each voiced-voiceless pair for each session; significance levels given are for one-tailed t-tests. Figures 1, 2, 3 and 4 plot the range of individual tokens of each stop consonant in each session of Tom, Tessa, Jane and Jay respectively. The table and figure for each child's data will appear in the appropriate section. The discussion in section 4.0 will focus on the similarities and differences across children.

In the following sections, reference will be made to primarily three categories of data (terms in quotation marks will be given precise definitions in the following paragraphs). CATEGORY I. Data show no evidence for the acquisition of a contrast: the tokens for both the voiced and voiceless stops fall within the "short lag region" (and typically show the same incidence of prevoicing); and the difference between the means for each pair of stops is not significantly different, nor is there a consistent relationship between the means such that over a number of sessions, the mean for the voiceless stop is always longer than the mean for the voiced stop. CATEGORY II. Data provide evidence for the acquisition of a contrast, which is, however, not adult-like: the difference between the mean VOT values of the voiced-voiceless stops is significant but the mean VOT values for both the voiced and voiceless stops fall within the adult perceptual "phoneme boundaries" for the voiced stop (Category II-A) or close to within these boundaries (Category II-B). Category II-A data are particularly interesting because the contrast that the child is making is presumably not perceptible to adults. For Category II-B data, adults would presumably have difficulty recognizing the contrast, since the majority of voiceless stop tokens typically fall within the adult perceptual boundaries for the voiced stop. CATEGORY III. Data resemble "adult values": the means for the child's voiced and voiceless stops fall within the appropriate phoneme boundaries; in Category III-A, the means for the voiceless stops (and usually the voiced stops also) are considerably longer than adult means, and in Category III-B, the means for the voiceless stops (and the voiced stops) are shortened back toward adult means. It appears that these categories may in fact reflect three general developmental stages, ordered as presented (see section 4.0).

We will adopt the following conventions for defining the phonetic regions on the VOT continuum. The "short lag region" will be from 0 to +20 for labial and alveolar stops (following Kewley-Port and Preston 1974 (apicals only); and Zlatin and Koenigsknecht 1976 (labials and apicals)), and from 0 to +40 ms for velars (cf. Zlatin and Koenigsknecht 1976). This definition of short lag as 0 to +20 ms has been adopted to facilitate direct comparison of our data to the data in Kewley-Port and Preston on which they base their claim that the VOT values for a majority of productions by children of the age we are studying will fall within the short lag region. In addition, as defined, it fits reasonably well with the Lisker and Abramson definition of short lag (0 to +25 ms) and with the range of short lag productions reported by those authors for English: /b/ 0 to +5 ms; /d/ 0 to +25 ms; /g/ 0 to +35 ms. We will need to refer to the "short lead region" for labials and alveolars; this region will be -20 to -1 ms (prevoiced velar stops were extremely rare). This short lead region, which is not precisely defined in any study, corresponds roughly to the "hole just below zero" that Lisker and Abramson state to be the one region on the VOT continuum which is not utilized in voicing systems of languages of the world. The "long lag region" will be +60 to +100 ms, and the "long lead region" will be -125 to -75 ms, following the Lisker and Abramson general descriptions (1964).

To arrive at some reference point for the adult language to which the children's data may be compared, we use the adult data reported in Lisker and Abramson 1964, Klatt 1975, and Zlatin 1974. Since these authors report slightly different mean VOT values for each stop, we will assume that a child's VOT mean for a stop is "adult-like" if it falls within the range of mean VOT values reported for that stop in the adult data: /b/ +1 to +11 ms; /d/ +5 to +17 ms; /g/ +21 to +27 ms; /p/ +47 to +65 ms; /t/ +67 to +75 ms; and /k/ +70 to +85 ms. Secondarily, the child's data will be judged adult-like if the

inter-place relationship is consistent (i.e. labials shorter than alveolars, which are shorter than velars) and the child's voiced and voiceless stops have non-overlapping ranges (a characteristic of adult stops reported in some but not all studies). Presumably, comparable adult data (e.g. fully conversational) would show overlap between ranges; however, the children must learn to produce at least fairly discrete categories, and to do so, they must reduce (if not eliminate) the number of extreme tokens. It was not expected that any of the children would show all these three characteristics of adult speech.

In discussing Category II-A data, we will find it helpful to discuss the mean VOT values for the children's production data in relation to the "phoneme boundaries" in adult perception. Similar to the slight disparity between figures in reports on the mean VOT values of stops produced by adult English speakers, there is also some disagreement as to the precise location of the boundary between the voiced and voiceless phonemes (i.e., the 50% perceptual cross-over point) in the perception literature on English. Here, we will adopt a compromise set of figures: +30 ms for labial and alveolar phonemes; and +50 ms for the velar phonemes (cf. Lisker and Abramson 1967b; Zlatin 1974; Pisoni and Lazarus 1974; and others). We do this in order to have a rough idea of the way in which an adult speaker of English would perceptually categorize the early productions of the children studied. Since this boundary is approximate, we will simply suggest that, when the means for both child phonemes (and the VOT values for the majority of tokens) fall within or nearly within the boundaries of one adult phoneme, an adult will either fail to perceive or at least have considerable trouble consistently hearing the contrast that the child is making. We will futher assume that an adult would categorically perceive the child's productions in most cases as belonging to the adult voiced phoneme, and suggest that, since the child continues to improve this contrast in the appropriate direction, it appears that the child does so in the absence of direct positive reinforcement from adults.

Yet we assume that speakers of any language may learn to modify their perceptual sets as an adjustment to the speech of unusual speakers (e.g. foreign speakers of English) or to unusual listening circumstances (e.g. as on a telephone line). It is possible that with repeated exposure to the child's speech, a parent could learn to hear the child's contrast. We will assume that there must be limits to such perceptual adjustment. For example, Hirsh 1959, Hirsh and Sherrick 1961, Stevens and Klatt 1974 and Pisoni 1977 claim that roughly 20 ms is the minimum amount of difference between the temporal onset of two events--such as between a burst and the onset of voicing--required for a listener to judge the two events as sequential, a judgment which would be necessary to the identification of English phonemic voicelessness. If approximately 20 ms does in fact represent the lower limit on the region of perceived temporal order sequence, it would be impossible for an adult (using only VOT as the perceptual cue) to hear the contrast that Tessa, for example, was making between /b/ and /p/ on session 6 when the VOT means were +5.07 ms and +14.43 ms respectively (*p=.029) and where 84% of all tokens fell within the 0 to +20 ms range. While the psycho-acoustic basis for the perception of voicing in initial position (e.g. the +20 ms argument in Pisoni 1977) may be disputed, the categorical nature of adult perception is not, and it is ultimately on this basis that we will assume there to be limits to the perceptual modification possible in natural language use situations.

In these cases where the children were achieving a significant difference between the means

and yet where both means fell within the perceptual boundaries for one adult phoneme (Category 2), we—as trained phoneticians—do not believe that we can reliably hear the contrast (except for those tokens that fell well outside the boundary), and we were in fact surprised by the results. Since VOT is only one of several acoustic cues for the perception of voicing in English, it is possible that some other attribute (e.g. burst amplitude or fundamental frequency of the following vowel, etc.) could be found in the children's productions and used by an adult to perceive the contrast; again, we can only repeat that we could not perceive the differences. Clearly, the perceptibility of these early 'within one adult phoneme' contrasts requires further investigation.[7]

To evaluate these early contrasts and the similar ('within one adult phoneme') cases where over a large number of sessions the mean VOT for the adult voiceless phoneme was consistently longer than and yet not significantly different from the mean for the voiced phoneme, we examined the variables of: (1) vowel height; (2) stress; (3) individual words; and (4) mode of elicitation. Both a high vowel and a stressed vowel (variables 1 and 2 respectively) have been reported by at least some researchers to lengthen the VOT of a preceding voiceless (i.e. long lag) stop and in some cases to affect the VOT of a preceding short lag stop (see Klatt 1975 and Smith 1975 on vowel height; and Lisker and Abramson 1967a and Klatt 1975 on stress). The lexical diffusion model of sound change (cf. Chen and Wang 1975) may be interpreted to predict that when the voicing contrast is beginning to emerge, long lag values would be found first in a few words and then gradually spread to other lexical items; thus, the difference in the mean VOT between a voiced and voiceless stop pair could be achieved by the values for a few individual words (variable 3). Barton 1976 found that for some children the voicing contrast was correctly produced in imitated utterances earlier than in spontaneously produced utterances (variable 4). In all cases, no correlation was found: thus it is not the case that the longer mean VOT values for the voiceless stops (in Category II) were caused by a greater incidence of voiceless stops in stressed and/or high vowel environments or in imitated productions, or by the behavior of individual words. Throughout the corpus for each child, tokens for any particular word showed the same range of VOT values as the general range of VOT values for the stop consonant as a whole (a few exceptions to the general pattern will be discussed in the following sections). Of the four variables, stress was largely irrelevant, since nearly all of the children's productions in these early stages were one word utterances of initial syllable stress stop words. We conclude then that such consistent differences (although small) between the mean VOT values for the voiceless and voiced stops indicate some attempt on the child's part to distinguish in production between the members of cognate pairs. These analyses were done for data in Category II and will be mentioned below where appropriate. Since no similar analyses of data in Category III were carried out, we therefore do not rule out the possibility that stress, for example, may have a significant effect on long lag stops.

In the following sections for each child, we will discuss the data for each child in relation to the three categories and in relation to the perception and production characteristics of adult voicing described above; the main goal will be to describe each child's individual path of development. In addition, each section will discuss other characteristics (in some cases, idiosyncratic) of each child's productions, as for example the distribution of initial stop words in the child's vocabulary ("lexical asymmetries").

3.1 Tom

At the beginning of the study when Tom was 1;6.24, his data revealed a bias towards words beginning with /b/, /d/ and /k/ (figures below are for the total number of different words and tokens produced on sessions 1-3): /b/ 12 words, 64 tokens; /p/ 6 words, 15 tokens; /d/ 7 words, 40 tokens; /t/ 3 words, 6 tokens; /g/ 1 word, 2 tokens; /k/ 7 words, 26 tokens. The distribution was, however, fairly even by session 7. Since not enough tokens of /t/ and /g/ were collected from the first anchor point (sessions 1 through 3), sessions 4 and 5 were also analyzed. Only the data from session 5 will be presented, since session 4 contained only 1 /g/ token and no /t/ tokens.

Tom's data show evidence for a significant voicing contrast at all three places of articulation throughout the study; although the mean VOT values for voiced and voiceless stops usually fell within the appropriate short and long lag regions, there were some changes over time.

On sessions 1-3 (1;6.24 to 1;7.7), the mean VOT values for /ptk/ are considerably shorter than adult mean values (Category II-B). The mean VOT values for /b/ and /d/ are within the range of adult /b/ and /d/ mean values (the mean for /b/ is +7.6 ms, excluding negative VOT tokens). Tom's two values for /g/ are similar to his means for /b/ and /d/ but are considerably shorter than adult /g/ values would be. (Table 3).

INSERT TABLE 3 AND FIGURE 1 HERE

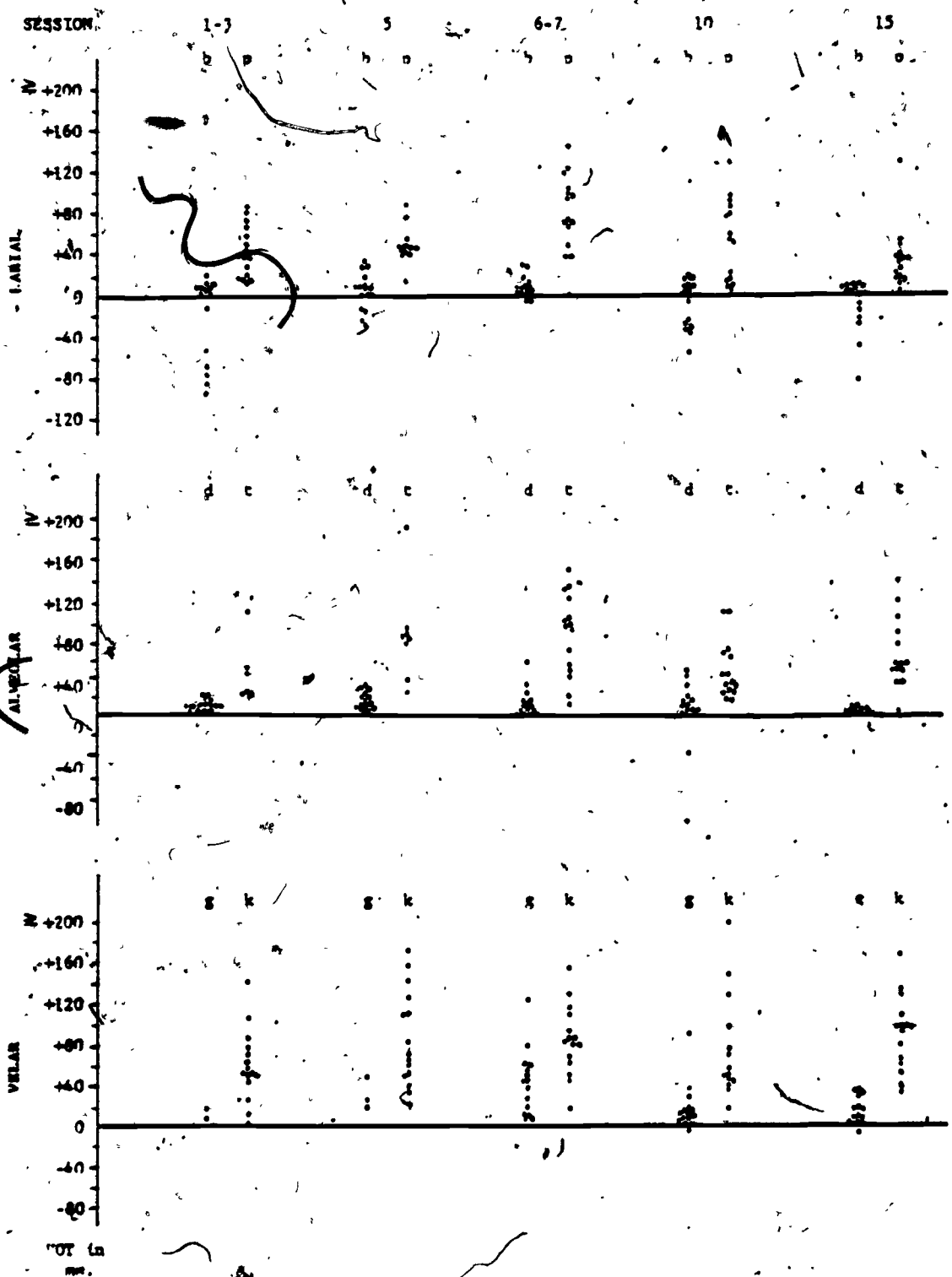
Although Tom's mean VOT values are not adult-like, there was very little overlap between the ranges for the voiced and voiceless stops, and Tom's contrasts were generally perceptible to adults. Only five /p/ tokens fell in the short lag region and six (40%) fell within the perceptual boundaries for the adult /b/ phoneme. For the alveolars, although there were only six /t/ tokens, there was no overlap between the /d/ and /t/ VOT ranges, and none of the /t/ values fell in the short lag region; however, three of the /t/ tokens (50%) fell within the adult perceptual boundaries for /d/. Similarly, for the velars, although there were only two /g/ tokens, both were short lag and only three of fifteen /k/ tokens fell within the short lag range; however, six /k/ tokens (40%) fell within the adult perceptual boundaries for /g/. (Figure 1). Thus, although the range of values for /p,t,k/ (up to +85 ms, +106 ms and +140 ms, respectively) show that Tom clearly had a phonological voicing distinction, the fact that 40% or 50% of all voiceless tokens fell within the adult voiced phoneme perceptual boundaries means that Tom had not fully mastered productive control over voicing, and it is likely that adults would label nearly half of his 'voiceless' stops as being phonemically 'voiced' (Category II-B); nearly half of his /p,t,k/ stops were in fact transcribed as voiceless, unaspirated.

Data from session 5 (1;8.29) are very different: the mean VOT values for /tk/ are considerably longer than adult mean values (Category III-A) and the mean VOT values for /bdg/ likewise increased (Table 3). The mean for /p/ falls at the lower end of the range (+47 to +65 ms) reported for /p/ in the speech of adults; since the mean for /p/ is within the adult range, it should be categorized as IIIB. This case is one of two (see section 3.2) that are exceptions to the general rule that the child's mean for a voiceless consonant shifts abruptly from Category IIB to Category IIIA (cf. the discussion in section 4).

Table 3. TOM: mean VOT values, number of tokens, standard deviations and ranges for each stop, by session.

Session	1-3	5	6-7	10	15
Age	1;6.24-1;7.7	1;8.29	1;9.12-1;9.23	1;11.8	2;1.17
/b/ mean	-14.27	+3.87	+7.60	-9.93	-11.07
N	15	15	15	15	15
S.D.	35.88	18.09	8.23	23.62	25.03
range	-87/+17	-28/+33	-4/+23	-56/+18	-80/+9
/p/ mean	+40.40	+48.58	+77.40	+36.87	+32.33
N	15	12	15	15	15
S.D.	26.17	16.50	38.40	33.20	29.96
range	+10/+85	+12/+67	0/+141	+8/+128	+3/+128
sig.level	p<.001	p<.001	p<.001	p<.001	p<.001
/d/ mean	+9.53	+15.47	+14.07	+5.00	+3.07
N	15	15	15	15	15
S.D.	6.34	8.93	15.71	35.47	3.63
range	+5/+21	+3/+31	0/+57	-102/+46	0/+10
/t/ mean	+45.67	+82.75	+60.53	+47.20	+61.93
N	6	8	15	15	14
S.D.	31.70	48.52	40.30	27.78	34.83
range	+24/+106	+24/+188	+12/+145	+17/+105	+3/+134
sig.level	p=.02	p<.001	p<.001	p<.001	p<.001
/g/ mean	+12.50	+30.00	+43.13	+17.47	+11.73
N	2	3	15	15	15
S.D.	7.78	15.13	32.50	23.38	12.98
range	+7/+18	+18/+47	+8/+126	-4/+96	-5/+34
/k/ mean	+59.53	+91.33	+85.27	+80.57	+91.60
N	15	15	15	14	15
S.D.	35.53	43.27	34.38	63.97	36.41
range	+2/+140	+35/+173	+17/+154	+15/+260	+34/+170
sig.level	p=.045	p=.015	p=.001	p=.002	p<.001

Figure 1. TOM: VOT values for each stop token, by place of articulation and by session.



Data from sessions 6 through 10 show that Tom is gradually shortening the VOT values for /tk/ and /d/ back toward adult-like values; the means for /b/ and /p/ increase on sessions 6-7 to values longer than adult values (Category IIIA) and then decrease on session 10 (Table 3). Unexpectedly, however, the mean VOT for /g/ continued to increase such that on sessions 6 and 7, it fell nearly in the perceptual boundaries for adult /k/ (n.b. Abramson and Lisker 1967b report approximately +40 ms as the boundary between /g/ and /k/ for adults). By session 15 (2;1.17)--the last recording session--, the means for most stops fall within the appropriate short or (moderately) long lag regions; however, his stops are not completely adult-like. For example, his mean for /p/ is considerably shorter than an adult /p/ would be, while the means for /t/ and /k/ are longer than the respective adult means would be. Without data from subsequent months, it is impossible to determine what trend is being established. Since studies of adults have shown that VOT values for stops embedded in sentences are shorter than the VOT values for stops in isolated words, it is reasonable for Tom's VOT values to decrease somewhat, since he was producing increasingly longer sentences as he was maturing; thus, the plausible hypothesis of the interaction of sentence length with stop VOT values would apply to the change seen in /p/ on session 15. However, the means for /t/ and /k/ (which had been decreasing since session 5) increased again on session 15; whether this represents what could be considered 'normal' fluctuation within a single stage or a significant developmental change cannot be known.

While there was the change in mean VOT values described above, it was nevertheless true that throughout the study, most VOT values for all six stops fell within the perceptual boundaries for the appropriate adult phonemes. For the voiced stops, Tom produced some tokens with voicing lead (range -104/-1); these prevoiced tokens were usually for the voiced labial stop. There were very few tokens of the voiced stops which fell in the adult voiceless phoneme boundaries. In contrast to the restricted range for voiced stop productions, the voiceless stops showed great variability: there was a wide range of values that extended well into the short lag region and well beyond the (adult) long lag region. While there are no published adult data collected under completely comparable conditions (i.e. fully conversational speech), it is probably that Tom's voiceless stops exhibited greater variability than adult data would. This wide range of variability was not due to individual lexical items: individual words showed the same range of VOT values as other words beginning with the same initial stop.

3.2 Tessa

On session 1 when Tessa was 1;4.28, she was using slightly more b-, d- and k-initial words than words beginning with p-, t- or g- (figures below are for the total number of different words and tokens produced on session 1): /b/ 8 words, 31 tokens; /p/ 3 words, 5 tokens; /d/ 3 words, 24 tokens; /t/ 2 words, 12 tokens; /g/ 4 words, 7 tokens; /k/ 5 words, 19 tokens.[8] Tessa spontaneously produced a large number of utterances (see Table 2), would freely and correctly respond to questions designed to elicit stop-initial words (and would also imitate such words), and in general showed no evidence of deliberate avoidance of particular word-types; this slight asymmetry quickly disappeared. The asymmetry that did exist at the beginning was not as obvious a factor in her acquisition of voicing as was the lexical asymmetry in Jane's data (see section 3.3).

Tessa produced lead voicing on /bp/ and /d/ on session 1: /b/ two tokens; /p/ one token; and /d/ five tokens. After this session, only /b/ was produced with lead voicing (range -34/-5 ms); the number of /b/ tokens produced with lead voicing varied from none (out of 15 tokens, sessions 8 and 16), 1 (out of 15, on both sessions 6 and 7), 2 (out of 14, session 11) to 3 (out of 15, session 3-4). Across all sessions and all stops, sixty percent of all prevoiced tokens fell in the short lead region (-20/-1 ms), and none were greater than -34 ms.

Tessa acquired a voicing contrast first at the alveolar, then at the labial and finally at the velar place of articulation; three stages reflecting this order are described below. In general, Tessa's data (more than Jane's) change over time moving from Category I, through Category III-A; this progress toward adult values is a gradual one, covering about a four month period.

3.2 Stage 1 (1;4.28 to 1;5.30)

Stage 1 (sessions 1 and 3-4) in Tessa's data is one in which she has a contrast at the alveolar (Category II-A) but not at the labial or velar places of articulation (Category I). On session 1 (1;4.28), the mean VOT for /d/ is significantly different from the mean for /t/ (Table 4). Further evidence for a contrast comes from the distribution of lead and long lag tokens: the only prevoiced tokens are for /d/, and the only tokens that fall outside the short lag region are productions of /t/ (Figure 2).

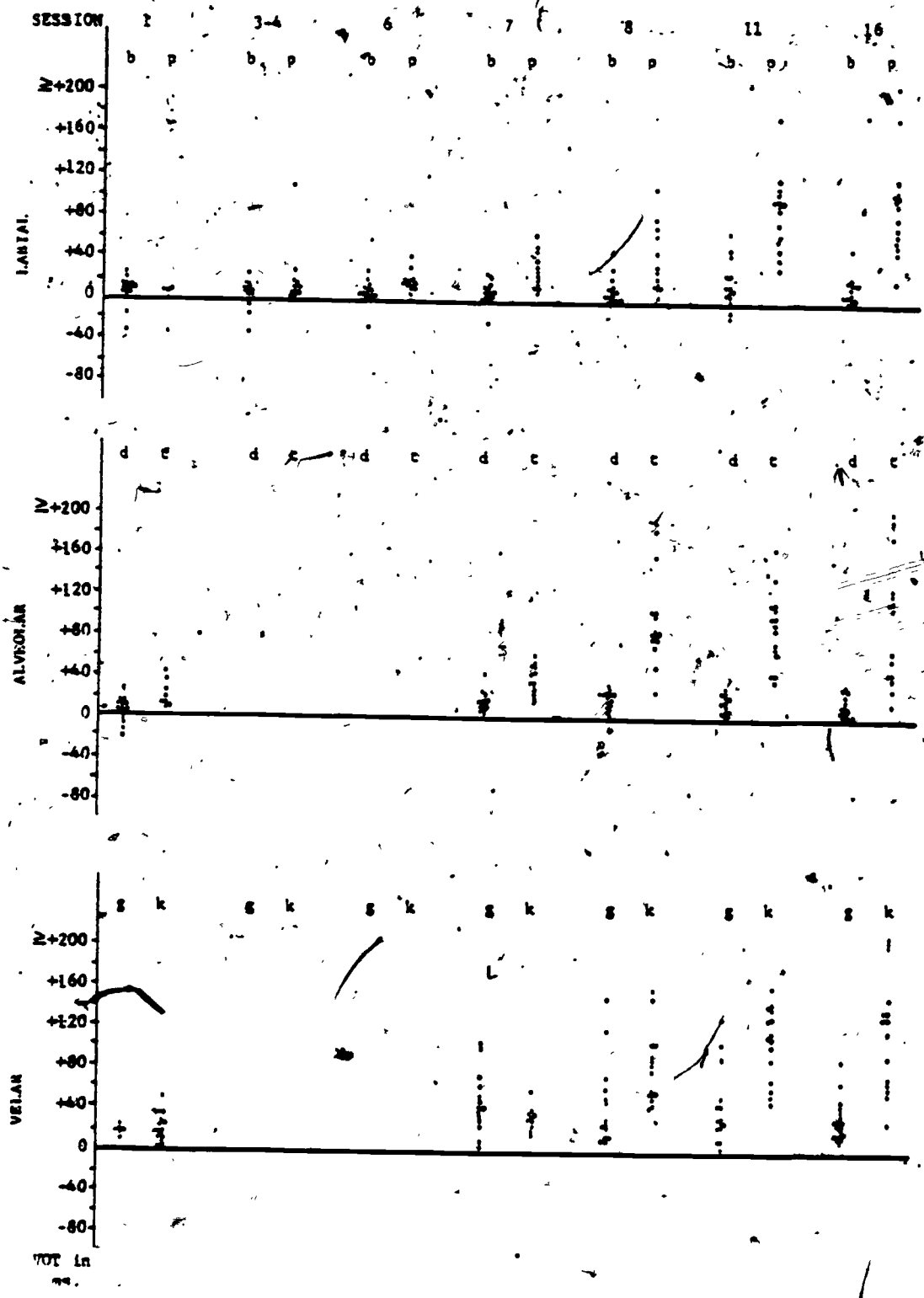
For the labials on session 1, the mean for /p/ is actually less than the mean for /b/, and the ranges for both are similar. On the sessions 3-4, the /p/ mean is longer than that for /b/. Neither difference is significant (Table 4). The difference in mean VOT between /b/ and /p/ on sessions 3-4 is largely caused by one production of pan, +111 ms. If we eliminate this token, the mean and range for /p/ becomes much more similar to those for /b/: /b/ +4.56 ms (range -34/+25 ms); /p/ +10.10 ms (range +4/+28 ms). This production of pan is the only one for this word in the corpus prior to session 11 (pan +74) at which point all /p/ tokens fall in the range +32 to +176 ms.

Pan is an exceptional utterance on this session in that no other words are produced with long lag voicing until session 8 (Figure 2). The long lag VOT for pan may indicate that this word was a lexical exception (on session 3-4) to Tessa's general rule for producing /p/ as short lag in which case it might indicate the beginning of a word-by-word spread of long lag voicing through her lexicon (cf. the lexical diffusion model). However, throughout Tessa's corpus, tokens for any particular word show the same range of VOT values as the general range of VOT values for the stop consonant as a whole: pants, for example, was produced as both short lag (+20 ms) and long lag (+67 ms and +75 ms) on session 8 when the range for p-words (0/+107 ms) covered the short to long lag region also. Since no other words were produced consistently as an exception to the general patterns for a given stage, we conclude then that the long lag production of pan was accidental or at least extra-systemic.

Table 4. TESSA: mean VOT values, number of tokens, standard deviations and range for each stop, by session.

Session	1	3-4	6	7	8	11	16
Age	1;4.28	1;5.18-1;5.30	1;7.6	1;7.13	1;7.27	1;9.2	2;0.10
/b/ mean	+4.33	+4.07	+5.07	+8.53	+11.20	+17.64	+12.93
N	15	15	15	15	15	14	15
S.D.	13.18	13.71	12.10	10.70	11.63	22.16	11.15
range	-32/+23	-34/+25	-27/+28	-22/+25	0/+44	-17/+65	+3/+49
/p/ mean	-1.80	+19.27	+14.43	+27.92	+34.93	+84.93	+93.40
N	5	11	14	12	14	15	15
S.D.	18.36	31.17	9.47	17.31	31.89	36.14	47.78
range	-34/+9	+4/+111	0/+40	+8/+60	0/+107	+32/+176	+19/+207
sig.level			p=.015	p<.001	p=.009	p<.001	p<.001
/d/ mean	+2.40			+15.53	+17.13	+16.23	+15.93
N	15			15	15	13	15
S.D.	10.81			9.61	8.48	8.38	8.96
range	-21/+14			+4/+42	+2/+32	+5/+31	+6/+32
/t/ mean	+20.50			+38.92	+98.33	+88.93	+100.20
N	8			13	15	15	15
S.D.	13.09			13.26	45.89	35.13	68.38
range	+7/+43			+19/+61	+27/+193	+38/+165	+13/+257
sig.level	p=.001			p<.001	p<.001	p<.001	p<.001
/g/ mean	+17.00			+47.00	+46.86	+43.33	+30.13
N	6			14	14	15	15
S.D.	3.63			28.59	41.31	35.46	21.42
range	+11/+22			+3/+100	+15/+150	+5/+126	+9/+85
/k/ mean	+19.93			+30.30	+79.60	+105.27	+125.93
N	15			10	15	15	15
S.D.	12.44			11.32	34.86	35.92	73.70
range	+3/+50			+17/+57	+43/+158	+46/+162	+26/+306
sig.level					p=.015	p<.001	p<.001

Figure 2. TESSA: VOT values for each stop token, by place of articulation and by session.



On session 1, there is no contrast between the velar voiced-voiceless stops: the difference between the mean for /g/ and that for /k/ is not significant (Table 4), and all /g/ tokens and all but one /k/ token fall within the short lag region (Figure 2).

3.22 Stage 2 (1;7.6 to 1;7.13)

During stage 2 (sessions 6 and 7), Tessa's data show evidence for a contrast between /b:p/ (Category II-A) and /d:t/ (Category II-B), but not for a contrast between the velar stops.[9] The mean VOT values and the majority of tokens for /b/ and /p/ on both sessions 6 and 7 fall within the perceptual boundaries for the adult voiced phoneme, as did the means for /d:t/ during stage 1. The difference between the mean VOT values for /d/ and /t/ (session 7) is also significant; note that the mean for /t/ falls just outside the perceptual boundaries for the adult voiced phoneme. (Table 4) On session 7, a majority of the tokens for both /p/ and /t/ fall outside the short lag region, while nearly all tokens of /b/ and /d/ are within this region (Figure 2). In separate analyses, no correlation was found between the length in VOT of the stops and whether the stop preceded a high vowel or occurred in any imitated production; all tokens were of initial stops in stressed syllables.

In contrast, the difference between the mean VOT values for /g/ and /k/ is not significant (Table 4), and the majority of tokens for /g/ (but not for /k/) fall for the first time outside the short lag region (cf. the distribution of velar tokens on session 1, Figure 2).

3.23 Stage 3 (1;7.27 to 1;9.2)

On session 8 (1;7.27), Tessa has a contrast between the voiced-voiceless stops at all three places of articulation: the data for /b:p/ fall in Category II-B, those for /d:t/ in Category III-A and those for /g:k/ in Category II-B (Table 4).[10] On session 11, all three places of articulation show highly significant differences also (Category III-A for all three places).

However, Tessa's stops are not adult-like. The mean VOT values for /b/ (sessions 6,7,8 and 11), for /d/ (sessions 7 and 8) and for /g/ (sessions 7 and 8) progressively increased, usually as the mean VOT values for the voiceless stops increased. As a result, Tessa's mean values for the voiced stops are more similar to adult mean values for these stops on session 1 than on session 8. Similarly, Tessa's voiceless stops also become less adult like over time: /p/ (session 11), /t/ (sessions 8 and 11) and /k/ (session 11) are considerably longer than the longest adult means, whereas previously the means for these stops had been gradually approaching the shortest of the adult means. The lack of correspondence between Tessa's voicing system and that of adults is also shown by the following: on session 8, her mean for /p/ falls nearly within the perceptual boundaries for adult /b/ (8/15 tokens ≤ 30 ms), and her /g/ mean falls nearly within the adult /k/ perceptual boundaries (5/14 tokens ≥ 50 ms).

The mean for /k/ on session 8 falls within the range (+70 to +85 ms) reported for /k/ in the speech of adults, and therefore should be categorized as IIIB. Tessa's /k/ mean on this session and Tom's /p/ mean on session 5 are the only cases which violate the general progression of data through Categories IIB to IIIA; in all other cases (i.e. Tessa's /p/ and /k/, Tom's /t/ and /k/ and Jane's /p,t,k/), the children's voiceless stop means are initially shorter than adult means (IIB) and then considerably longer than adult means (IIIA). Thus, in seven out of nine cases, the stage-wise nature of IIB followed by IIIA holds; in the remaining two cases, a transition stage occurs (chronologically) between IIB and IIIA data--a stage in which the child's voiceless stop mean is in fact adult-like (i.e. a stage technically "IIIB" in the present descriptive system). The surprising near-absence of such 'transition' data will be discussed in section 4.0.

An examination of the range of productions in Figure 3 shows that Tessa's /p/ tokens gradually move into the adult /p/ range (sessions 1 through 11) and that her /t/ and /k/ somewhat more quickly move into the appropriate adult ranges. For all six consonants, the ranges of values are considerably larger than adult ranges for the same stops; it appears that Tessa is in some sense exploring phonetic space in her attempt to determine the appropriate phonetic targets (see section 4.0).

Stage 3.24 Stage 4 (2;0.10)

For Tessa as for all the children, the final stage in the development is one in which adult-like productions are achieved. For Tessa, the VOT mean and range for all six stops (i.e., Category III-A, session 11) must be shortened back toward adult values.

However on session 16 (2;0.10)--the last time she was recorded--, the mean VOT values for /ptk/ are longer than on session 11 (Table 4). These means, which are considerably longer than adult means would be, have increased as the ranges of tokens have increased (Figure 2). In contrast, the mean VOT values for /bdg/ have decreased, as compared with the respective means on session 11 (Table 4). The mean for /d/ is within the adult range of mean VOT values (+5 ms to +17 ms). The shortening of the mean for /g/ is important: on session 16, the /g/ mean is well within the adult perceptual boundaries for /g/ for the first time since session 1.

3.3 Jane

Tabulation of all stop-initial words produced on each session shows that Jane's sessions 1 through 7 (1;6.19 to 1;8.24) were characterized by a preponderance of b-initial words (16 words/130 tokens). Words beginning with p- and g- were comparatively uncommon: during the same time period, she used only three p-initial words (16 tokens) and only three g-initial words (3 tokens); k-initial words were more common (6 words and 23 tokens). On session 8 when she was 1;9.15, she was using roughly equal numbers of all stop words except for g-initial words; at this point she still only used

three such words. By the next session (when she was 1;10.7), she was using at least eight words of each type.

In contrast to the lexical asymmetry at the labial and velar places of articulation, the number and frequency of d- and t-initial words were roughly equal over time. Although at the beginning, d- and t-initial words were few in number (four words of each type) (and d-initial words were produced slightly more often), from sessions 1 through 7, she produced these words frequently: five d-initial words (33 tokens) and nine t-initial words (24 tokens).

The above pattern of vocabulary acquisition correlated in an interesting way with Jane's acquisition of the voicing contrast. She first had a contrast of voicing at the alveolar place of articulation, and the number and frequency of d- and t-initial words were approximately equal from the beginning. She acquired the contrast between /b/ and /p/ on session 8--a development which was accompanied by a large increase in the number and frequency of p-initial words; prior to session 8, the number and frequency of b-initial words was significantly greater than that for p-initial words, and the VOT values for /p/ nearly all fell in the adult /b/ range as did the VOT values for /b/. On session 9, when the number and frequency of g-initial words increased substantially, the difference between the mean VOT of /g/ and /k/ is significant for the first time; prior to this point (from sessions 1 to 8), Jane used three g-initial words (only 6 tokens) and seven k-initial words (38 tokens), and /g/ was treated by Jane as if it were /k/ (cf. session 8 discussed below where all /g/ tokens are moved to the long lag region at the same time that /ptk/ shift).

Among the four subjects, Jane used the greatest amount of voicing lead. Throughout the study, 9% to 40% of all /b/ tokens were produced with voicing lead. From sessions 1 through 8, a similar proportion of /p/ tokens was produced with voicing lead (14% to 29%), but the range for lead was considerably narrower (-35 to -6 ms) than the range of lead for /b/ (-139 to -2 ms). Jane stopped producing /p/ with voicing lead after session 8, the session on which the /b:p/ contrast was acquired. She began producing voicing lead on /d/ at sessions 4 and 5 and on /g/ on session 11; /t/ and /k/ were never produced with voicing lead. The consistency with which she produced voicing lead on /b/, the use of long lead (in addition to short lead) and the possible spread of voicing lead to /d/ and /g/ suggest that Jane may be one of the set of English speakers that produce at least some proportion of the voiced stops with voicing lead.

[INSERT TABLE 5 AND FIGURE 3 ABOUT HERE]

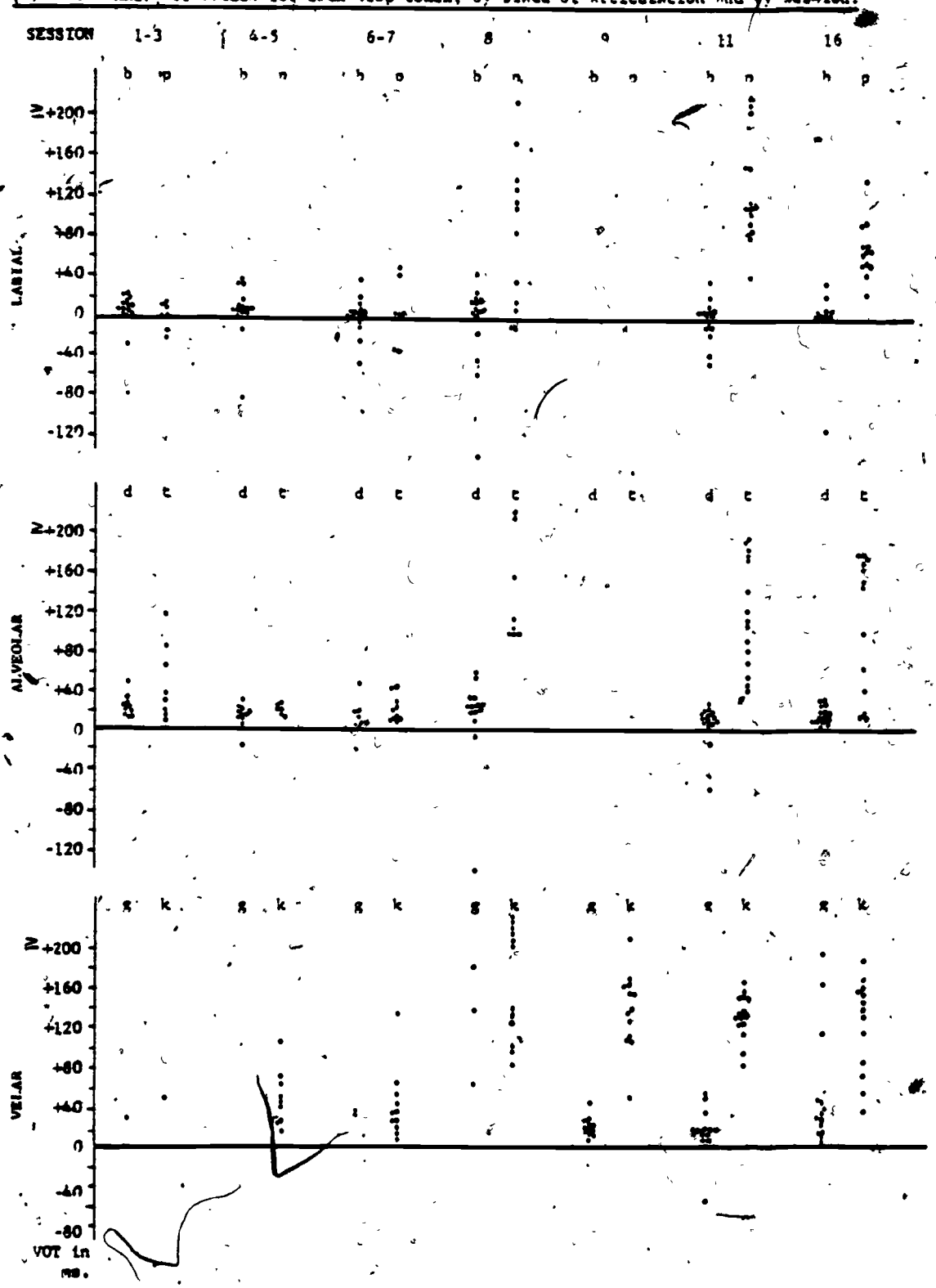
As previously mentioned, Jane (like Tessa) acquired the voicing contrast first at the alveolar, then at the labial and finally at the velar place of articulation; this developmental order will be described as three stages in the following discussion of her data. Unlike Tessa however, only one set of data (/d:t/, stage 1) fit Category II-A, and none fit into Category II-B. In contrast to the gradual progression of voiceless stop tokens into the long lag region seen in Tessa's data, the same transition in Jane's data appears to be much more abrupt; however, this 'abruptness' may be partially due to sampling interval.

3.31 Stage 1 (1;6.19 to 1;8.24)

Table 5. JANE: mean VOT values, number of tokens, standard deviations and ranges for each stop, by session.

Session	1-3	4-5	6-7	1-7	8	9	11	16
Age	1;6.19-1;6.29	1;7.12-1;7.26	1;8.20-1;8.24	1;6.19-1;8.24	1;9.15	1;10.7	1;11.0	2;1.16
/b/ mean	+2.73	+7.00	+2.00	+3.91	-6.67		-0.60	-0.91
N	15	15	15	45	15		15	11
S.D.	26.54	27.91	18.54	24.21	43.95		20.62	37.81
range	-81/+24	-81/+39	-45/+39	-81/+39	-139/+43		-46/+38	-110/+86
/p/ mean	-0.29	-	+5.57	+2.64	+90.50		+128.87	+71.39
N	7	0	7	14	12		15	13
S.D.	13.02	-	33.00	24.29	87.55		60.45	26.91
range	-21/+14	-	-35/+51	-35/+51	-7/+289		+41/+248	+25/+136
sig.level					p=.002		p<.001	p<.001
/d/ mean	+24.70	+12.82	+14.25	+17.31	+15.80		+7.67	+17.20
N	10	11	8	29	15		15	15
S.D.	11.59	10.98	15.44	13.26	44.85		19.17	9.03
range	+10/+48	-13/+30	+1/+48	-13/+48	-135/+60		-35/+25	+4/+34
/t/ mean	+47.13	+20.83	+22.30	+30.21	+151.38		+122.13	+109.23
N	8	6	10	24	8		15	13
S.D.	39.77	5.64	13.50	26.63	73.06		54.59	68.76
range	+8/+118	+14/+28	+8/+45	+8/+118	+100/+270		+44/+197	+13/+178
sig.level				p=.019	p<.001		p<.001	p<.001
/g/ mean	+29.00	-	+38.00	+35.00	+131.67	+22.08	+14.47	+54.14
N	1	0	2	3	3	12	15	14
S.D.	-	-	1.41	5.29	59.18	9.85	23.52	59.97
range	+29	-	+37/+39	+29/+39	+68/+185	+7/+45	-49/+54	+3/+195
/k/ mean	+53.00	+51.56	+41.25	+46.00	+177.42	+144.77	+134.27	+124.62
N	1	9	12	22	12	13	15	13
S.D.	-	29.16	33.89	30.89	69.62	49.78	21.29	46.87
range	+53	+16/+110	+6/+134	+6/+134	+85/+285	+52/+274	+88/+166	+35/+190
sig.level						p<.001	p<.001	p=.001

Figure 3. JANE: VOT values for each stop token, by place of articulation and by session.



The first stage in Jane's data is one in which her productions show some evidence for a contrast at the alveolar but not at the labial or velar places of articulation. On sessions 1-3, 4-5 and 6-7, the mean for /t/ is always at least 8 ms longer than the mean for /d/ (Table 5). On sessions 1-3 and 6-7, the distribution of tokens for /t/ is different from that for /d/. In the data from combined sessions 1-7, the difference between the means is significant (Table 5); in addition to the one-tailed t-test, a two-way analysis of variance was done which was also significant (*p=.018). Note that the mean for /t/ falls almost within the adult perceptual boundaries for the voiced phoneme (Table 5), and that most tokens of /t/ fall within the adult perceptual boundaries for /d/ (Figure 3). In separate analyses, we examined the VOT for all tokens in relationship to vowel height, stress and the variable spontaneous-imitated; no correlation was found. However, a few productions of two words were unusual.

On session 1 (1;6.19), there were three /t/ tokens that were produced with long lag: toy +86 ms, +118 ms; and teddy +68 ms; these tokens are unusual in that no other /t/ tokens (nor any /d/ tokens) were produced with long lag until session 8, at which point all /t/ tokens (and no /d/ tokens) were produced with long lag. The two productions of toy and the one of teddy could be interpreted as evidence that Jane was correctly using long lag on at least these two words (cf. the lexical diffusion model). This explanation seems unlikely for the following reasons: toy was produced three times on the same session with short to moderately long lag voicing (+8 ms, +11 ms and +39 ms); and one production for a voiced stop (/b/) was produced with long lag (session 4-5). In addition, the ranges for both /d/ and /t/ included a substantial proportion of tokens in the moderately long lag region also: 35% of the /d/ tokens fall between +21 and +48 ms, and 46% of the /t/ tokens fall between +21 and +45 ms (sessions 1-7). As in the case of pan in Tessa's corpus, we conclude that the production of these long lag stops approximately 3 months prior to the long lag stage does not indicate a significant lexical parameter at the phonetic level of VOT.

At stage 1 (sessions 1-7), there is no evidence for a contrast between the labial or velar phoneme pairs: (1) the means for both /b/ and /p/ on sessions 1-3, 6-7 and 1-7 fall between 0 and +10 ms (well within the short lag region); the combined 1-7 sessions show that /b/ and /p/ are nearly identical (cf. the means and standard deviations, Table 5); and both /b/ and /p/ tokens show nearly the same incidence of prevoicing (Figure 3); (2) the means for both /g/ and /k/ fall within the range of 0 to +52 ms (Table 5), and the three /g/ tokens fall within the range for /k/ tokens (+6 to +134 ms, Figure 3).

3.32 Stage 2 (1;9.15)

Stage 2 (session 8) is marked by the acquisition of a contrast between /b:p/, a highly significant difference between /d/ and /t/, the production of /ptk/ as consistently long lag stops, and additional evidence for a lack of a contrast between /g:k/. On first glance, it appears that the shift of /ptk/ to the long lag region occurs very abruptly. However, session 8 was conducted three weeks after session 7 (due to the child's illness). We assume then that sampling interval is at least partially responsible for the appearance of sudden change (cf. Tessa's session 7 and 8 data which were collected at a two week interval (Figure 2) and which show a less dramatic change).

Between session 7 and 8, Jane began producing at least some /p/ tokens with long lag voicing. In the production of p-initial words for session 8 (1;9.15), we find that the only short lag values occur in productions of "old words" (i.e. those words which had been in her productive vocabulary since 1;6.19). These words are pop +37 (once also +289 ms), pig -6 ms, +9 ms, +18 ms (once +137 ms) and pumpkin -7 ms. All other p-initial words produced on this session were "new" (i.e. produced for the first time) and all were produced with long lag voicing (+84/+173 ms). This distribution of short lag tokens by word type (i.e. in old words only) could be interpreted as evidence for the gradual spread of long lag through the lexicon. However, each of the new words was produced only once; with more tokens, these words may also have shown variability.

By this same session, the VOT values for /t/ and /k/ also move well into the long lag region, and the means for all voiceless stops are considerably longer than adult means (Category III-A). The difference between the means for /t/ and /d/ is significant, and the ranges are non-overlapping. In contrast, the difference between the means for /k/ and /g/ is not significant; as Jane moved /k/ into the long lag region, she similarly shifted /g/. The two words effected are: girl which had been produced with a VOT of +29 ms on session 2 was produced as +68 ms on session 8; goat which had been produced as +39 ms on session 7 was produced as +142 ms and +185 ms on session 8; goes (+37 ms on session 7) did not occur in the corpus for session 8.

3.33 Stage 3 (1;10.7)

On session 9, Jane's productions for /g/ and /k/ show a significant difference (Category III-B), and the ranges do not overlap. This was accomplished by a shift of /g/ back into the short lag region.[11] Whether this development constitutes perceptual learning, an advance in production or a reorganization at the phonological level cannot be known. To posit phonological organization as the source, one would need to assume a priori (and in the absence of evidence) that Jane perceived the phonological and phonetic contrast between /g/ and /k/ in the adult language and, secondly, that a general phonological constraint on the complexity of her own system allowed only a two-way voicing contrast (/b:p/ and /d:t/). The context-free phonological rules needed would be: Stage 1, /g/→[g] and /k/→[g]; Stage 2, /g/→[k], /k/→[k]; and Stage 3, /g/→[g], /k/→[k] (where, in all three stages, [g] is voiceless, unaspirated and [k] is voiceless, aspirated). The motivation for such a rule change is unclear (given the assumption that the child's perception is not a factor), and we know of no cases in the literature where a child has a contrast, produces one member of the pair correctly (/g/, Stage 1) and then loses the correct production (Stage 2), except in those cases where a context-sensitive rule has been added to the grammar. We conclude then that it seems likely that the change in /g:k/ at stage 3 indicates that Jane has perceptually learned to contrast /g/ and /k/. If the explanation were articulatory in nature, it would be difficult to explain why she initially moved /g/ from the short lag to the long lag region (session 8).

To shed additional light on this issue, we return to the pattern of vocabulary acquisition discussed at the beginning of this section. On session 8, when Jane shows the first evidence of having acquired /b:p/, the number of p-initial words in her productive vocabulary increased from three to

eight. Three weeks later, on session 9, when she shows the first evidence of having acquired /g:k/, the number of g-initial words also increased from three to eight. In contrast, the number and frequency of d-initial words were similar to the number and frequency of t-initial words throughout sessions 1 to 7 (where the difference between the means, when sessions are collapsed, is significant), and both d- and t-initial words occurred more frequently in the corpus than either p- or g-initial words. It should also be noted that not only the total number of individual p- and g-initial words increased on session 8 and 9 respectively, but the number of tokens for both /p/ and /g/ similarly increased. At least two explanations are possible: either Jane did not acquire the contrasts between /b:p/ and /g:k/ until session 8 and 9 because she did not 'know' sufficient p- and g-initial words to recognize the contrast; or she perceived the differences between /b:p/ and /g:k/ sufficiently well to 'avoid' p- and g-initial words, because she knew she could not pronounce these sounds (cf. the phonologically motivated avoidance strategies in Ferguson and Farwell 1975). At least in the case of /g:k/, the former explanation seems most likely.

3.34 Stage 4 (1;11.0 to 2;1.16)

Stage 4 (sessions 11 and 16) which had not been completed by the end of our study, involves the mastery of adult-like VOT values for each stop. Jane was making progress toward this goal by shortening the VOT values of /ptk/ and /bd/ (Category III-A). In contrast, the mean for /g/ unexpectedly increased such that it once again fell within the adult perceptual boundaries for /k/. The long mean for /g/ (+54.14 ms) is caused by three values: goose +164 ms; and garbage +116 ms and +195 ms. All other values are ≤46 ms. On the same session goose was also produced as +27 ms; on the previous session analyzed (session 11), goose was produced as +13 ms and garbage as +6 ms. Thus, this change in the mean for /g/ (on session 16) was not caused by the idiosyncratic behavior of a newly acquired word but is rather further evidence of the variability with which individual words were produced.

3.4 Jay

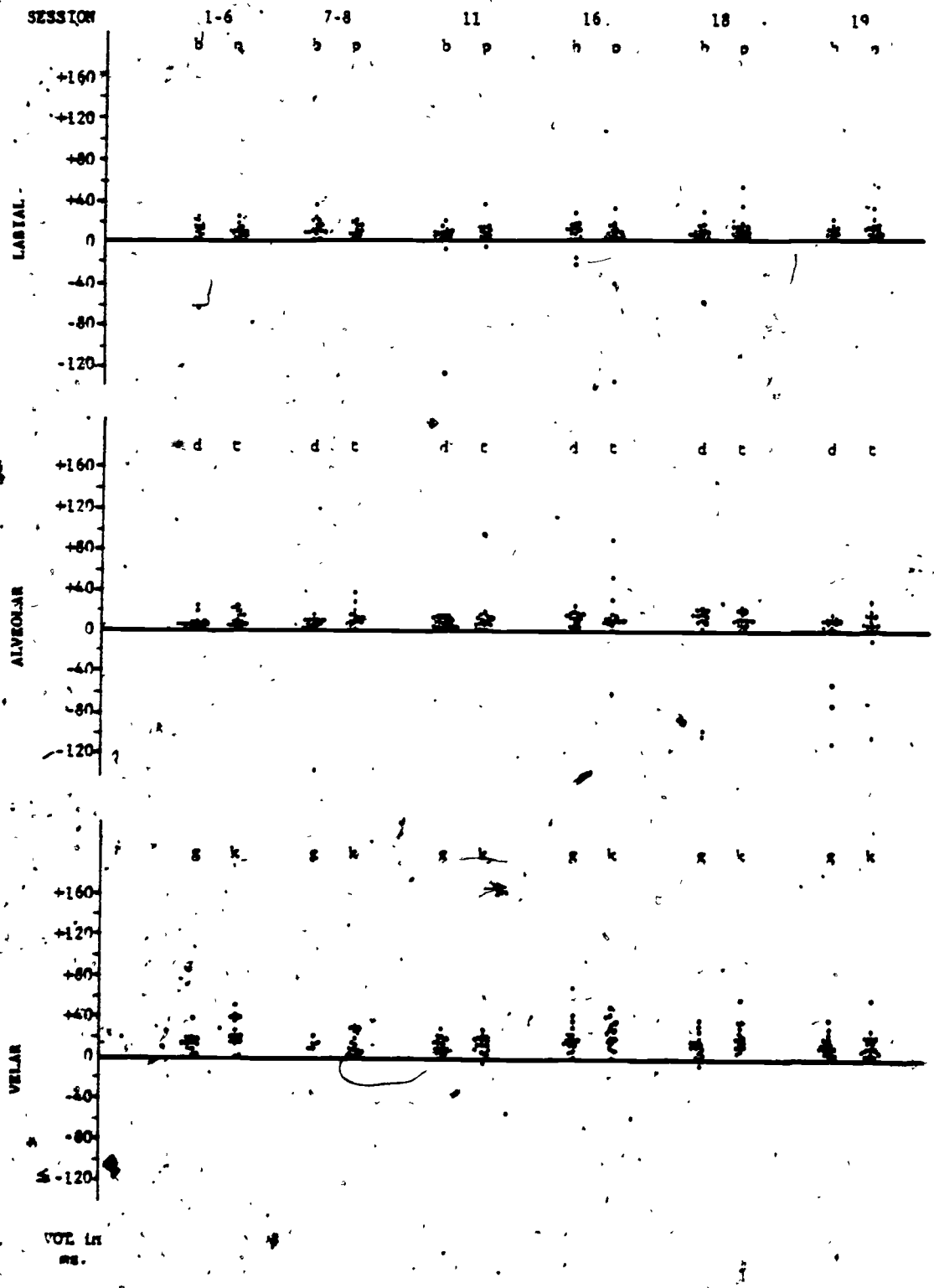
Between the ages of 1;7 and 2;4 Jay did not acquire a voicing distinction in initial position that was discernible to the adult ear. Almost all initial stops were transcribed as voiceless and unaspirated; there were very few errors of place and these were usually attributable to assimilation. At the beginning of recording, there was a marked lexical asymmetry in his data such that words beginning with b-, d- and k- predominated (figures below are for the total number of different words and tokens produced on sessions 1-3): /b/ 12 words, 44 tokens; /p/ 3 words, 8 tokens; /d/ 6 words, 20 tokens; /t/ 2 words, 5 tokens; /g/ 3 words, 4 tokens; /k/ 8 words, 29 tokens. By session 9, the distribution was fairly even.

Jay's data were analyzed in six anchor points. These results are given in Table 6 and Figure 4. Throughout the study the VOT distributions had basically the same characteristics. Since there was no apparent change over time, the data for each stop consonant across all sessions were collapsed, and

Table 6. JAY: mean VOT values, number of tokens, standard deviations and ranges for each stop, by session.

Session	1-6	7-8	11	16	18	19	1-19
Age	1;7.9-1;9.22	1;10.5-1;10.15	2;0.0	2;2.2	2;3.0	2;4.2	1;7.9-2;4.2
/b/ mean	+4.13	+11.60	-2.87	+5.27	+4.87	+5.50	+4.71
N	15	15	15	15	15	10	85
S.D.	20.26	9.30	33.81	11.17	18.47	5.30	19.31
range	-65/+24	0/+35	-123/+20	-19/+25	-55/+31	0/+17	-123/+35
/p/ mean	+9.07	+9.55	+7.82	-3.00	+13.20	+8.13	+7.34
N	15	11	11	15	15	15	82
S.D.	6.73	6.19	10.97	39.27	18.71	9.18	19.20
range	0/+24	+2/+20	-5/+38	-134/+33	+1/+54	0/+33	-134/+54
sig. level							
/d/ mean	+6.13	+7.47	+7.53	+10.20	-4.85	-11.20	+2.72
N	15	15	15	15	13	15	88
S.D.	6.32	3.50	2.53	6.99	42.07	36.46	23.20
range	0/+23	0/+14	+3/+11	0/+25	-104/+19	-108/+13	-108/+25
/t/ mean	+10.50	+16.60	+14.47	+14.47	+9.93	+0.13	+11.03
N	14	15	15	15	14	15	88
S.D.	7.43	8.70	23.22	31.90	7.13	29.63	21.21
range	0/+23	0/+37	+1/+96	-62/+92	+2/+23	-102/+28	-102/+96
sig. level	p=.05	p<.001					p=.007
/g/ mean	+15.50	+12.80	+13.00	+22.20	+14.27	+15.27	+15.86
N	12	5	15	15	15	15	77
S.D.	9.00	4.55	7.62	17.03	11.06	8.88	11.17
range	+4/+38	+9/+20	0/+27	+2/+71	-3/+35	+3/+33	-3/+71
/k/ mean	+25.40	+15.80	+15.40	+24.13	+22.87	+14.33	+19.66
N	15	15	15	15	15	15	90
S.D.	14.41	9.76	7.73	12.94	11.07	12.78	12.22
range	0/+55	+3/+31	-2/+26	0/+48	+6/+47	+3/+53	-2/+55
sig. level	p=.025				p=.021		p=.02

Figure 4. JAY: VOT values for each stop token, by place of articulation and by session.



these results are also given in Table 6. It will be seen that the majority of VOT values for voiced and voiceless stops fell within the short lag range. Taking short lag to be 0 to +20 ms for labials and alveolars and 0 to +40 ms for velars, then 89.4% of all tokens fell within the short lag range. Over all three places of articulation, 94.8% of tokens fell within the range 0 to +40 ms. Of the remaining tokens, 1.9% had VOT values greater than +40 ms and 3.3% were produced with voicing lead. While most of the tokens in the range >20 ms occurred in productions for an adult voiceless phoneme, some were for an adult voiced phoneme; similarly, while most of the tokens produced with voicing lead occurred in words which began with an adult voiced phoneme, some were for the adult voiceless phoneme. The lead tokens were spread throughout the range from -1 to -140 ms.

INSERT TABLE 6 AND FIGURE 4 ABOUT HERE

Although Jay did not seem to acquire a voicing distinction, there were some anchor points on which the differences between the mean VOT values corresponding to adult voiced-voiceless stops were significant. As Figure 4 shows, the significance on these sessions could not be attributed to outlying VOT values. The data can best be discussed by looking at the three places of articulation separately.

LABIALS. There was never a significant difference between the mean VOT values for /b:p/. In four of the six anchor points, the difference was in the expected direction (i.e. the mean for /p/ is greater than the mean for /b/), but in the remaining two anchor points, the difference was in the opposite direction. In the data from the combined sessions 1-19, the mean for /b/ was not significantly different from the mean for /p/ (Table 6). Since Jay's labial data show no evidence for a contrast, they fit into Category 1. There were nine examples of voicing lead (six for /b/, three for /p/ Figure 4), and nine examples of continuous voicing (all but one in /b/ tokens). There were more tokens produced with lead and continuous voicing towards the end of the study.

ALVEOLARS. On all the anchor points, the mean for /t/ was longer than the mean for /d/; the difference between the means was significant on two occasions (sessions 1-6 and sessions 7-8, Category II-A). Furthermore, when sessions 1-19 were combined, there was a significant difference between the mean VOT values (Table 6); in addition to the one-tailed t-test, a two-way analysis of variance was carried out and was also significant ($p=.011$). There were eight examples of voicing lead: 2 tokens of /d/ on session 18; 3 /d/ tokens on session 19; 1 /t/ token on session 16; and 2 /t/ tokens on session 19 (Figure 4). There were also thirteen examples of continuous voicing: 1 /d/ token on session 18; and 12 /d/ tokens on session 19.

VELARS. As with the alveolars, the relationship between the means for the voiced and voiceless stops was such that the mean VOT value for the voiceless stop was always longer than the mean for the voiced stop, except on session 19 where the two means were nearly identical. The difference between the means was significant on two anchor points (sessions 1-6 and session 18). For the combined sessions 1-19, both the one-tailed t-test (Table 6) and a two-way analysis of variance ($p=.021$) showed the difference between the /g/ and /k/ mean VOT values to be significant (Category II-A). There were two examples of voicing lead: 1 /k/ token on session 11 and 1 /g/ token on session 18. There were five instances of continuous voicing: 1 /g/ token and 4 /k/ tokens, all on session 19.

For both alveolars and velars then, overall there was a significant difference between the means for the voiced and voiceless stops; nevertheless there was a high degree of overlap between the ranges, and most tokens of both voiced and voiceless stops fell within the short lag regions. Before concluding that Jay was attempting to produce a voicing distinction between /d:t/ and /g:k/ but was unable to reproduce an adult-like contrast, we investigated several other variables that potentially could have affected the results. We looked at the possible effect of vowel height, stress, individual lexical items, and mode of elicitation. None of these variables appeared to show significant differences.

We also considered the possibility that some systematic experimenter bias could have influenced the measurement of voiced and voiceless tokens. However, this seemed unlikely for two reasons. First, when we began the study we did not expect the children to have a voicing contrast in their speech and we especially did not expect to find consistent differences in VOT values for short lag productions of the voiced and voiceless stops; however, the differences found in alveolars and in velars were present from the beginning. The second reason why experimenter bias is an unlikely explanation is that there was never a significant difference between the labial voiced and voiceless stops; it is rather implausible that such a bias would affect places of articulation differentially.[12]

We therefore conclude: (1) that Jay was attempting to produce a difference between the voiced and voiceless phonemes at at least two places of articulation (Category II-A); (2) that the contrast he was maintaining fell within the adult voiced phoneme boundaries and thus would not be perceptible to adult speakers; and (3) that without instrumental analysis of a large number of utterances, no evidence could have been found to indicate that he had a phonological voicing contrast at these two places of articulation.

Unlike the other children who as early as 1;5 or 1;7 began to produce some adult-like voiceless stops, Jay had not acquired an adult-like voicing contrast by the end of the study, when he was 2;4.2. Thus, he will be at least 11 months older than our youngest child when he begins producing long lag voicing for /ptk/. In other aspects of language development Jay was comparable to the other children: in the total number of different words he used and the number of tokens he produced per session (see Table 2) he was similar to the other children (especially to Jane). To compare the children to each other on general language development we roughly computed MLU (mean length of utterance), following Brown 1973.[13]

All four children had a MLU of 1.0 on their respective first sessions; for Tessa this rose above 1.0 on her third session (age 1;5.18); for Jane on her second session (1;6.22); for Tom on his fourth session (1;7.21); and for Jay on his fifth session (1;9.4). Tom therefore had an MLU of 1.0 when he had a highly significant voicing distinction at all three places of articulation (session 1). Tessa had an MLU of about 1.93 when she had these distinctions (session 8), and Jane had an MLU of about 1.57 when she had these distinctions (session 9). On his last full session, Jay had an MLU of 2.23: it seems then that his general language development (as measured by MLU) at the end of the study was ahead of the general language development of the other children at the stages at which they acquired roughly adult-like voicing contrasts.

Clearly, it is not the case that MLU correlates directly with the acquisition of voicing: the data from Tom, Tessa and Jane show that MLU may vary from 1.0 to 1.93 at the time at which a fairly adult-like voicing contrast is produced at all three places of articulation (cf. Bond and Wilson 1977). However, we present MLU data to show that Jay was progressing normally with respect to general language development; it is also true that in other phonological aspects he was progressing normally (e.g. he used at least one fricative (in medial and final positions), a postvocalic /r/, several final position consonants clusters, and many poly-syllabic words). Since the age range reported in the literature for the acquisition of the voicing contrast (1;4 to 2;8) is also quite large, we conclude that it is possible that Jay is within the range of individual differences which can be considered normal for the acquisition of voicing in English.[14]

4.0 Discussion

In this section, the data will be reviewed in order to characterize the general stages (4.1) and to provide a brief summary of other major findings (4.4). In addition, the data will be examined for evidence which shed light on the nature of the skills being acquired (4.2) and the process of change within each child's system as related to two major models for acquisition (4.3).

To facilitate comparison of the four subjects, some data from section 3.0 are repeated here in Figure 5. This figure plots the change over time in mean VOT values for each stop for each child. For reference points, the adult perceptual phoneme boundaries and the adult mean VOT values for each stop are also included in the figure. To avoid unnecessarily complicating the figure, only one mean VOT value for each adult stop is given; these mean VOT values are from Lisker and Abramson 1964 (but note that the Lisker and Abramson means for the voiced phonemes exclude negative VOT values, whereas the means given for the children, in some cases, include negative values). For each child, the VOT means are presented separately by place and in left-to-right order for labial, alveolar and velar stop pairs. Beneath each set of means are the session(s) from which the means come and the category in which they fall. In this section, the discussion will refer to data presented in Figure 5 unless otherwise noted.

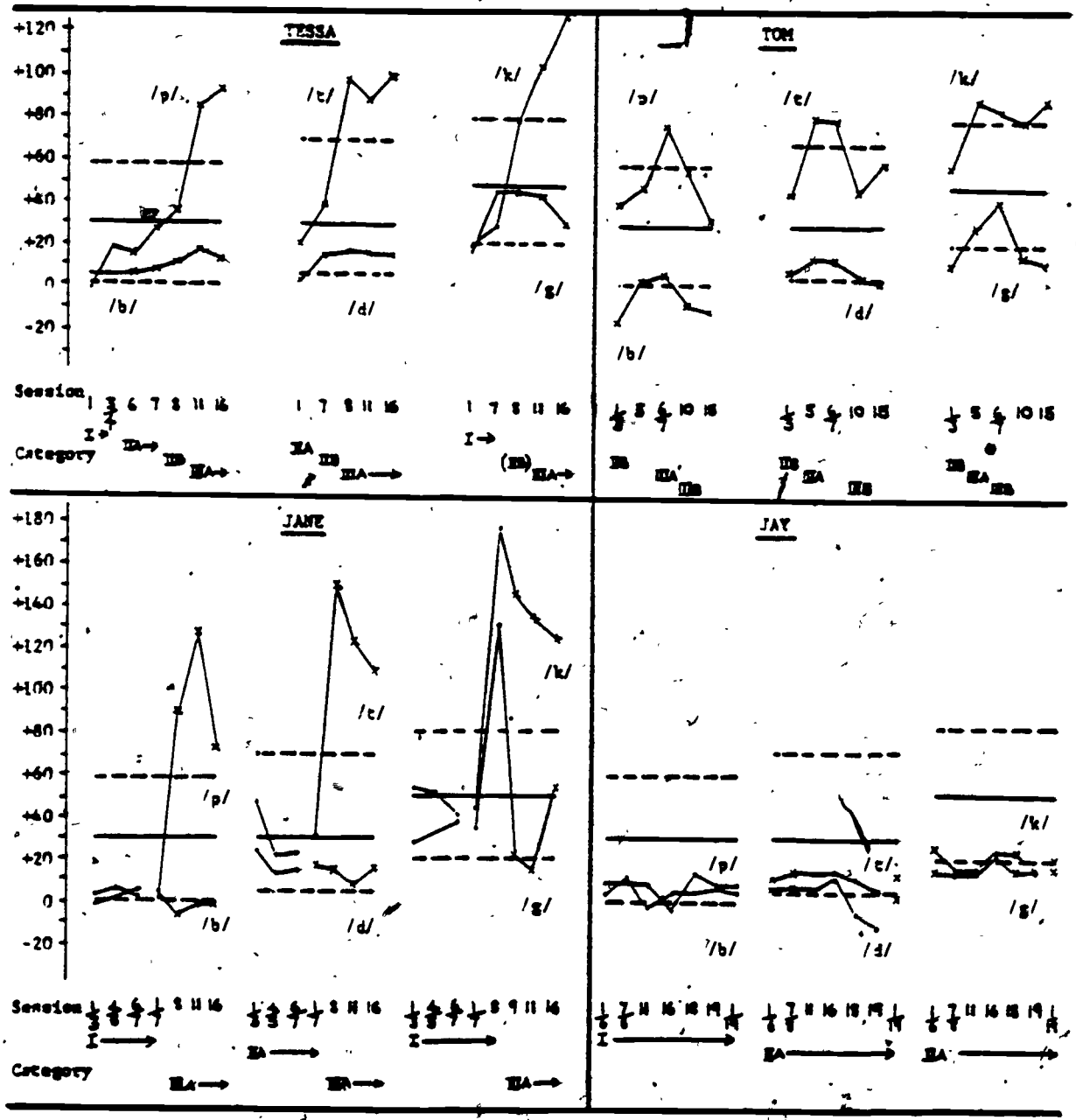
[INSERT FIGURE 5 ABOUT HERE]

4.1 Stages to the acquisition of voicing as seen in these data

In general, the pattern which emerges from the data is one in which there are three stages to the acquisition of voicing: (I) the child has no contrast (Category I); (II) the child has a contrast but one which falls within or nearly within the perceptual boundaries of one adult phoneme (Categories II-A and II-B); (III) the child has a contrast which resembles the adult contrast (Categories III-A and III-B).

STAGE I. In stage I, the child produces predominately short lag stops (cf. also Kewley-Port and Preston 1974). Short voicing lead will be produced occasionally for either adult voiced or voiceless phonemes. Similarly, long lag voicing if produced at all appears in productions of either adult voiced or voiceless stops.

Figure 5. Changes in VOT means over time, for each stop, by child.



STAGE II. In stage II, the child is still producing most stops with short lag voicing; however, for the first time, the mean VOT for the adult voiceless stop is consistently longer across several points in time than the mean VOT for the adult voiced stop. Depending on the child, the resulting difference between the means is either consistently significant (Tessa) or significant on some sessions but not on others (Jane and Jay). Children also differ in how they change the voiceless stops during this stage: Tessa gradually lengthened the VOT of her voiceless stops (and thereby increased the contrast between the voiced and voiceless stops), while Jane and Jay did not. During this stage, the child typically produces short voicing lead and continuous voicing on adult voiced stops only, and may begin to produce some adult voiceless stops with (moderately) long lag voicing.

The stage II(A) data are interesting in that the contrast that the child is maintaining was not noted by the transcribers and is presumably not perceptible to adults in general. Dean and Huntington 1976 report a similar case: when the phonemically voiceless stop productions of esophagal speakers were examined spectrographically, a statistically significant number were distinct (in the appropriate VOT direction) from their productions of phonemically voiced stops; listeners, however, predominantly labelled all productions as 'voiced'. In five separate dialect studies, Labov et al. 1972 found that speakers, as for example in Tillingham (Essex), consistently made small differences in their own speech which served to maintain the identity of word classes, but that these same speakers could not label these differences on conscious reflection (in minimal pairs and in commutation tests), either in their own speech or in the speech of others from the same dialect area. These authors conclude that "intuitive judgments of 'same' or 'different' are not necessarily a reliable base on which to build a theory of phonological development" (Labov et al. 1972, 254). The 'phonological development' referred to here is historical change, but the comment equally applies, particularly in the present context, to phonological change in children. The results of the present study clearly show that, in at least some cases, the judgments of adults may not capture significant facts about the child's system and that spectrographic analysis can provide insight in precisely those areas where adult perception fails.

STAGE III. At the beginning of stage III, the child is producing adult voiceless stops with extremely long lag voicing: the mean VOT values for the voiceless stops are considerably longer than adult means would be. Moslin 1976 (reporting on two children approximately 1;4) and Gilbert 1977 (reporting on data from six children ages 2;7 to 3;3) also found that children's voiceless stop means are longer than adult means would be. Zlatin and Koenigsknecht 1976 report that two-year-olds (ages 2;6 to 3;0) and six-year-olds (ages 6;1 to 6;11) produce shorter mean VOT values for voiceless stops than do adults.

In general, as the means for the voiceless stops change, the means for the voiced stops also change in the same direction, thus frequently becoming less like adult voiced stop means than in earlier stages; this tandem relationship holds true for most changes in stage II also. Thus, the children appear to be in some sense exploring phonetic space, possibly because they are not sure of the precise target for the voiced and voiceless stops (although they have a relative notion of the appropriate distance between the two members of a pair) or possibly simply because they cannot control their productions. Moslin and Nigro 1976 report that for /t/ in the speech of adult-to-child, the VOT values are

considerably longer than the VOT values (for /t/) in the speech of adult-to-adult. Thus, our subjects may be attempting to match the longer VOT values that occur in the input they hear. Since, however, Moslin and Nigro also report shorter VOT's for /d/ in adult-to-child as compared to adult-to-adult discourse, we would need some other explanation (e.g. exploring phonetic space) to explain why our subjects increase the VOT values for the adult voiced phonemes. In contrast to the Moslin and Nigro 1976 results, Baran, Zlatin Laufer and Daniloff 1977 found that for their three adult female speakers there were, overall, no significant differences between adult-directed and child-directed conversation; Baran et al. found significant differences only between citational forms and conversational speech.

In this stage, the child's voiceless stop distributions are wide and flat, with extreme values at times exceeding +250 ms. The VOT values for the voiced stops generally cluster in the short lag region, with a few values in both the short and the long lead regions, and the range generally overlaps with the range of VOT values for the corresponding voiceless stop.

Later in stage III, the child begins to shorten the VOT for both the voiced and voiceless stops back toward the adult values: extreme values are produced less often. None of the children consistently achieved non-overlapping ranges, although the partial clustering of VOT values provides evidence for the improved discreteness of the voicing categories. Overall, the labial stops are shorter than the alveolar stops, which in turn are shorter than the velar stops. Only one child (Tom) was producing all three stop pairs in a fairly adult-like manner.

The transition between stages is in general a gradual one: over a two to four month period (1;7 to 1;9, Tom; 1;5 to 1;9, Tessa; and 1;7 to 1;10 Jane), the child learns to produce the voiceless stops consistently with long lag voicing. However, the transition from stage II to stage III occurs fairly abruptly: the month and a half interval between sessions 3 and 5 shows more change in Tom's data than any subsequent similar period and similarly for the three week interval between Jane's sessions 7 and 8 compared to any preceding or subsequent three week interval (Figure 5). Tessa's alveolar data show the discontinuity between stages II and III most clearly: in the ten weeks between sessions 1 and 7, the mean for /t/ increased 18 ms, while in two weeks between sessions 7 and 8 (stages II and III respectively), the mean for /t/ increased 59 ms (Figure 5).

4.2 Hypotheses on the skill being acquired

At issue here then is whether the changes in VOT seen in these data reflect the development of a single homogeneous skill or the interaction of two (or more) phonetic skills. In the first case, the child may be viewed as achieving a contrast by progressively lengthening the VOT of the voiceless stops; in the second view, the child would be achieving a contrast by first using one phonetic component (stage II) and subsequently adding a second phonetic component (stage III), both of which result in differences in VOT. The case for skill interaction is strengthened by the following: for the most part (i.e. in seven out of nine cases), there is no transition stage (i.e. between stages IIB and IIIA) in which the mean VOT values for the voiceless stops fall within the range for adult mean VOT values

for /ptk/ (cf. sections 3.1 and 3.2). If we assume the children to be using a simple algorithm 'lengthen VOT for the voiceless stops', then we have the unusual situation in which an entire range of the VOT continuum has been skipped--the range which is moreover the precise range of adult values.

If the child is uniformly attempting to lengthen the VOT of the voiceless stops, it would be reasonable to assume that the abruptness of the change between stages II and III, and thus the appearance of discontinuity, are largely caused by sampling interval. It may be that at the point at which the child begins to produce voiceless stops with long lag voicing, changes occur fairly rapidly in the child's production; in which case, a two week recording interval (and obviously a three and four week interval also) is too great a time interval to show the full curve of progressively longer mean VOT values. Certainly, the development of many motor skills show faster skill improvement at some time periods than at others. Similarly, the acquisition of many motor skills show learning curves similar to that shown in these data: a stage showing a gradual increase toward the target (stage II) followed by a stage in which the target is over-shot (stage III-A).

However, it may be that a discontinuity does exist in the children's development such that stages II and III represent the acquisition of different underlying skills. In this view, stage III could correspond to the child's acquisition of the aspiration component of adult voiceless stops, whereas during the previous stage, the child was using a different phonetic component to distinguish adult voiced-voiceless stops. Leopold (1947) says that his daughter acquired the voiced-voiceless contrast at 2;0. His daughter's contrast was not adult-like however: in Hildegard's speech, the adult 'voicing' contrast was marked by the presence (in voiceless stops) or absence (in voiced stops) of "energetic release". In our data, late stage II (since early stage II is presumably not perceptible) may be similar to the "energetic release" stage described by Leopold; if this were the case, the children in stage II would be separating stops on the basis of a feature like 'tenseness' (or burst amplitude). Another possibility would be that stage II data represent the children's attempt to reproduce solely the first formant differences between adult voiced and voiceless stops (note that the mean VOT values for the voiceless stops in stage II (A and B) data range from +10 ms to +40 ms, the approximate area of F1 cutback in adult voiceless stops).

The crucial difference between the two hypotheses is the following: the single skill hypothesis assumes that the child recognizes the temporal differences between voiced and voiceless stops and attempts to achieve this contrast by progressively lengthening the VOT of the voiceless stops; the skill interaction hypothesis assumes that the child has broken the voicing contrast down into two (or more) fairly distinct perceptual components and that in stage one the child is attempting to match one component and in stage two changes his or her productions to incorporate the second component. The first hypothesis would predict that changes in VOT over time would show a progression through the complete VOT continuum, whereas the second hypothesis would predict a qualitative break in the changes in VOT at some point on the continuum. Since the virtual absence of transition data in the range for adult /ptk/ means may be due to sampling interval, and since changes could occur fairly abruptly in either the single skill or the skill interaction hypotheses, we have no way to conclusively resolve this issue. It should be obvious, however, that although VOT may be thought of

as a single continuum, it does in fact represent the conflation of several acoustic and articulatory components; thus, changes in VOT do not necessarily indicate the development of a single underlying skill.

4.3 Two models for developmental sound changes

The speed with which the children move from stage II into stage III may partially account for the 'across-the-board' model of change discussed by Smith 1973. This model, which is similar in some ways to the nineteenth century, neogrammarian model for historical change, claims that when a phonological change occurs in a child's system, it does so rapidly and applies to all relevant forms (cf. the neogrammarian tenet that sound changes have no exceptions). In contrast, the 'lexical diffusion' model, which was originally used to describe historical changes in Chinese, claims that 'a phonological change propagates itself gradually across the lexicon, from morpheme to morpheme (Chen and Wang 1975, 255)'. In studies of child phonology acquisition, evidence in support of the latter model has been presented by Hsieh 1972 and by Ferguson and Farwell 1975. The two models differ primarily in predictions about both the rate of change (rapid versus gradual) and the nature of the change (few if any lexical exceptions versus many lexical exceptions). Both models as used in the child phonology literature would predict at least some variability during the stage at which a change is taking place.

Although the children's phonetic acquisition of adult-like long lag voicing does appear to occur quickly (stage III-A), we have evidence that the children are making a phonological contrast even earlier, that at least one child (Tessa) was improving this contrast throughout stage II, and that for the two children for whom we see the acquisition process completely (Tessa and Jane), the spread of the voicing contrast across all three places of articulation takes approximately three months. Since at some point, the children's voiceless stops move into the perceptual boundaries for the adult voiceless stop (stage II-B), an adult's judgment of 'across-the-board' change is more directly a function of the categorical nature of the adult's perception than a function of any categorical change in the children's production; thus, to the extent that the children lengthen the VOT of the voiceless stops (from stage II-A through stage II-B), the description 'across-the-board'--which implies rapid phonological change--is erroneous. Since the 'across-the-board' label could be applied to the changes seen at stage IIIA, the model accurately describes a phonetic change but cannot account for the slow (three month) phonological changes in the stop system as a whole; the gradualness of this change is predicted correctly by the lexical diffusion model.

The data from all four subjects overwhelmingly demonstrate that the phonetic VOT range for any phoneme will be found in any word which begins with that phoneme, provided it is produced a number of times. Thus, in general, we have little evidence for the lexical parameter assumed by the lexical diffusion model; insofar as this lexical homogeneity refutes this model, we have evidence for a model which would predict the lexical exceptionless nature of the change, as the across-the-board model would. Sampling, however, is again a critical factor. Although we are confident that our data adequately represent each child's vocabulary (see section 2.2), we clearly did not record, on each session, all the words that the children knew; in addition, it is possible that we did not record at

frequent enough intervals to fully document the time of change (cf. the preceding discussion). Thus, it may be that our method was ill-suited to provide data of the sort required by the lexical diffusion model.

In summary, the generally lexical-exceptionless nature of the developmental changes is correctly predicted by the across-the-board model, and the gradual spread of the phonological voicing contrast through the stop system corresponds to the concept of gradual change found in the lexical diffusion model. However, some words do appear to be exceptions to the pattern in a particular stage (e.g. gan, Tessa; toy and teddy, Jane), and in one child's data long lag voicing may have first appeared in 'new' words; these data could be evidence for the lexical diffusion model. Similarly, the transition into stage IIIA could be described as 'across- the-board'. Thus, neither model is completely correct but both have important elements that correspond well with particular aspects of the data reported here.

4.4 Additional findings

In addition to evidence for three stages in the acquisition of voicing in the stop system as a whole, we have some data on the order of acquisition within the system and data on the effect of place of articulation on the process. For two of the children (Tessa and Jane), the voicing contrast appears first at the alveolar, then at the labial and finally at the velar place of articulation; these children also establish non-overlapping ranges first at the alveolar place (see sections 3.2 and 3.3). For one child (Jay), evidence for a contrast appears in the alveolar and velar data but not at all in the labial data. For three children (Tom, Tessa and Jane), voicing at the velar place presented more difficulty than at the other two places. For both Tom and Tessa, /g/ shifted from the short lag region into the longer lag production range and thus almost into the adult perceptual boundaries for /k/; Jane went through a short stage when /g/ was produced with extremely long lag values. For all four children, voicing lead was produced significantly more often at the labial place of articulation (see sections 3.1-3.4), and in general labial mean VOT values were shorter than alveolar means, with velar means the longest. Clearly, place of articulation, with the associated aerodynamic and physiological differences, is an important variable in the production of voicing. Possibly related to the differential affect of place of articulation on the production of voicing is the lexical asymmetry found in the vocabularies of all four children in the earliest sessions. For at least one child (Jane), this asymmetry—in particular the low frequency of words beginning with /p/ and /g/ in her vocabulary—appeared to be an important factor in the acquisition of the voicing contrast itself.

The range of individual differences with respect to age of acquisition is striking. Three children acquired a relatively adult-like voicing contrast at all three places of articulation by approximately 1;9; the fourth child who was in all other respects developing normally had not done so by 2;4.[15] More importantly, the contrast that this child was maintaining between /d:t/ and /g:k/ most likely could not have been detected by parents or other adults. Similarly, the earliest contrasts of two other children (at approximately 1;5 Tessa and 1;8 Jane) and possibly the first contrast of the third other child (at least by 1;7 Tom) presumably would not have been reliably detected by adults.

In conclusion then, the voicing contrast appears to be learned very early: evidence that children have acquired the appropriate phonological contrast may be found in the productions of children as young as 1;5. However, it may take up to eleven months before the children's productions improve to the point where the contrasts that the children are making may be perceived by adults. Considerable progress toward the production of an adult-like voicing contrast is typically made by the age of 2;0, although there are striking individual differences. Although considerable progress is usually made by 2;0, it may be many months (or even years) before children acquire sufficient articulatory skill to consistently produce adult-like voicing (Zlatin and Koenigsknecht 1976; Gilbert 1977).

Notes

1. This research is part of the activities of the Stanford Child Phonology Project and has been supported by a National Science Foundation Grant (BNS 76-08968) to Charles A. Ferguson and Dorothy A. Huntington, Departments of Linguistics and Hearing and Speech Sciences, Stanford University. We gratefully acknowledge their support during all phases of the research. We would also like to thank Harold Clumeck, John Kingston and Deborah Ohsiek for their assistance at various stages of the data collection and Lise Menn, Carl Muller and Marsha Zlatin Laufer for comments on an earlier version of this paper. A summary of this study was given October 1, 1977, at the Second Annual Boston University Conference on Child Language Development, and a preliminary version of this paper appears in the Stanford working paper series Papers and Reports on Child Language Development (1977) #14.

2. This study is part of a cross-linguistic study of the acquisition of the voicing contrast in initial stops in English, Spanish and Cantonese. Data from the other languages and cross-linguistic comparisons will be treated in subsequent publications of the Project.

3. In general, children appear to learn the voicing distinction among stops in initial position before in other positions in the word; however, the children in Velten 1943, Smith 1973 and Menn 1971 apparently learned the distinction in final position first.

4. To insure that this study would have longitudinal data from four children, a fifth child was recorded also. Had one of the first four withdrawn from the study, the fifth child would have become one of the four subjects. Since the first four stayed in the study, data from the fifth child were not instrumentally analyzed.

5. We are grateful to Linda Watson for conducting this session and to Jay and his parents for consenting to this additional visit.

6. Although consonant harmony is a widespread phenomenon in child phonology (see Vihman 1977), our subjects, who were evidently beyond the consonant harmony stage, rarely made place errors (less than 1% of the productions showed place errors). Nevertheless, the 'correct place of articulation' criterion was required to insure that the English data would be comparable to those collected from the Project Spanish and Cantonese subjects, who produced more harmonized forms: in all three studies then only 'correct' stops were analyzed. Although we know of no study which demonstrates that, during the stage when voicing is not contrastive, a stop substitute in a harmonized production (e.g. the initial [k] in [kokiz] for doggie) is any different from other stops of the same place (e.g. [k] in [kokət] for goat or the [k] in [kitl] for little), the possibility that such phones are different could not be discounted a priori; this is a topic that deserves separate investigation.

7. A preliminary study on this issue is planned: phonetically trained students (unfamiliar with the results of the present report) will be asked to transcribe approximately 75 stops characterized by very short VOT values (i.e. ≤34 ms); the stimulus items to be transcribed are a randomly selected subset of Category [illegible] data from Jay.

8. Tessa also produced 10 tokens of doggie on this session. However, since this word was produced with an initial velar stop, it has not been counted in the tabulations. This was the only assimilated word in her initial stop word corpus. Throughout the study, she produced doggie with an initial velar stop.

9. Tessa's data may have shown a significant difference between /b:p/ prior to sessions 6 and 7; the tape for session 5 was accidentally destroyed. On sessions 3 and 4, the difference between the means for /b/ and /p/ was not significant.

10. The data for /g:k/ are put in Category II-B, not because the mean for the voiceless stop falls nearly within the adult voiced boundary (as in the cases of other II-B data), but because the mean for /g/ falls nearly within the adult /k/ boundary.

11. As in the shift of /ptk/ in to the long lag region, this shift of /g/ into the short lag region appears to be abrupt. However, session 9 was conducted three weeks after session 8 (again because of the child's illnesses), and we assume that the apparent 'sudden change' is largely due to sampling interval.

12. To investigate this further, we submitted a set of spectrograms (with gloss unmarked) to a third observer for independent measurements. Although the VOT values for specific items frequently differed from the original measurements, the magnitude of the difference between the means for each voiced-voiceless pair was the same in the re-measured and original sets. This agreement demonstrates that the original observers had not been influenced by knowing the gloss.

13. Note, however, that our experimental procedures, which encouraged one word utterances, will tend to give a lower MLU.

14. We visited Jay in Boston in October 1977, when he was 2;6.19. At that time, he was producing all voiceless stops with strong aspiration. Although we may reasonably assume that the VOT characteristics of his productions were not identical to those for adults (i.e. probably longer VOT values than would be found in adult speech), he clearly had acquired an acceptably adult-like voicing contrast at all three places of articulation by 2;6.19--i.e. at least one month earlier than the oldest child reported in the literature. This confirms our earlier hypothesis that he was within the normal range.

15. The fourth child acquired an adult-like contrast by at least 2;6.19 (see section 3.4). On the basis of the transcriptions, the fifth child recorded had acquired an adult-like contrast at all three places of articulation by at least 1;6.27, her first recording session.

References

Baran, Jane A., Marsha Zlatin Laufer and Ray Daniloff. (1977). Phonological contrastivity in conversation: a comparative study of voice onset time. J Phonetics, 5.339-350.

Barton, David. (1976). The role of perception in the acquisition of phonology. Ph.D. dissertation. University of London.

Bond, Z.S. and H. F. Wilson. (1977). Voicing in the speech of language-delayed children. Paper presented at the 93rd meeting of the Acoustical Society of America (June).

Brown, Roger. (1973). A First Language: The Early Stages. Cambridge, Mass.: Harvard University Press.

Burling, R. (1959). Language development of a Garo and English speaking child. Word, 15(1).45-68.

Bush, C.N., M.L. Edwards, J.M. Luckau, C.M. Stoel, M.A. Macken, and J.D. Petersen. (1973). On specifying a system for transcribing consonants in child language: a working paper with examples from American English and Mexican Spanish. Stanford University: Department of Linguistics.

Chen, Matthew and William S-Y Wang. (1975). Sound change: actuation and implementation. Lg. 51.255-281.

Chomsky, Noam and Morris Halle. (1968). The Sound Pattern of English. New York: Harper and Row.

Dean, C. Richard and Dorothy A. Huntington. (1976). The acoustical bases of the voiced-voiceless distinction in esophegal speech. Paper presented at the annual meeting of the American Speech and Hearing Assoc. (November). Houston, Texas.

Ferguson, Charles A. (1977). New directions in phonological theory: language acquisition and universals research. In Roger W. Cole (ed.), Current Issues in Linguistic Theory. Bloomington: Indiana University Press.

_____ and Carol B. Farwell. (1975). Words and sounds in early language acquisition. Lg. 51(2).419-439.

Gamkrelidze, T. V. (1975). On the correlation of stops and fricatives in a phonological system. Lingua. 35.231-262.

Gilbert, John H.V. (1975). A voice onset time analysis of apical stop production in three-year-olds. J.Ch.Lg. 4.103-110.

Hirsh, I. J. (1959). Auditory perception of temporal order. JASA. 31:759-767.

_____, and C. E. Sherrick. (1961). Perceived order in different sense modalities. J.Exp.Psychol. 62.423-432.

Hsieh, Hsin-I. (1972). Lexical diffusion: evidence from child language acquisition. Glossa. 6.89-104.

Huntington, Dorothy A., Harold Clumeck, Marlys Macken and Deborah Ohsiek. (1977). Some methodological considerations on the study of VOT in children. ms. Stanford University: Hearing and Speech Sciences.

Jakobson, Roman. (1941/1968). Child Language, Aphasia and Phonological Universals. The Hague: Mouton. (Translated by A. Keiler, 1968.)

Kewley-Port, Diane and Malcolm S. Preston. (1974). Early apical stop production: a voice onset time analysis. J. Phonetics. 2.195-210.

Klatt, Dennis H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. JSHR. 18.686-706.

Labov, William, Malcah Yaeger and Richard Steiner. (1972). A Quantitative Study of Sound Change in Progress. Volume I. Philadelphia: U.S. Regional Survey.

Ladefoged, Peter. (1971). Preliminaries to Linguistic Phonetics. Chicago: The University of Chicago Press.

Leopold, Werner F. (1947). Speech Development of a Bilingual Child. A Linguist's Record. Volume II: Sound-Learning of the Two-Year Old Child. New York: AMS Press.

Lin, Sin-Chih. (1971). Phonetic development of Chinese infants. Acta Psychologica Taiwanica. 13.191-195.

Lisker, Leigh and Arthur S. Abramson. (1964). A cross-language study of voicing in initial stops: acoustical measurements. Word. 20.384-422.

_____ and _____. (1967a). Some effects of context on VOT in English stops. Language and Speech. 10(3).1-28.

_____ and _____. (1967b). The voicing dimension: some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, Prague: Academic Publishing House of the Czechoslovak Academy of Sciences (1970).563-567.

Macken, Marlys A. (1976). Permitted complexity in phonological development: one child's acquisition of Spanish consonants. Papers and Reports on Child Language Development. 11.28-60. (To appear Lingua).

Major, Roy C. (1977). Phonological differentiation of a bilingual child. Ohio State University. Working Papers in Linguistics. 22.88-122.

Menn, Lise. (1971). Phonotactic rules in beginning speech. Lingua. 26.225-251.

Moslin, Barbara J. (1976). Development of the voiced-voiceless contrast in English stops: a VOT analysis of two mother-child dyads. Paper presented to the 7th meeting of the Northeast Linguistic Society. (To appear NELS 7).

_____ and Georgia Nigro. (1976). Apical stop production of mothers to their children at 10 months and 16 months: a VOT analysis. Paper presented at the 1st Annual Boston University Conference on Child Language Development.

Pisoni, David B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. JASA. 61(5).1352-1361.

_____ and Joan H. Lazarus. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. JASA. 55(2).328-333.

Smith, Bruce. (1975). Effects of recent context and sex on VOT in initial voiced labial stop consonants in English: preliminary observations. Texas Linguistic Forum. 2.152-160.

Smith, Neilson V. (1973). The Acquisition of Phonology, A Case Study. Cambridge: University Press.

Srivastava, G. P. (1974). A child's acquisition of Hindi consonants. Indian Linguistics. 35.112-118.

Stevens, Kenneth N. and Dennis H. Klatt. (1974). The role of formant transitions in the voiced-voiceless distinction for stops. JASA. 55(3).653-659.

Velten, H. V. (1943). The growth of phonemic and lexical patterns in infant language. Lg. 19.281-292.

Vihman, Marilyn M. (1978). Consonants harmony: its scope and function in child language. In J.P. Greenberg, C.A. Ferguson and E.A. Moravcsik (eds.), Universals of Human Language. Stanford: Stanford University Press.

Zlatin, Marsha A. (1974). Voicing contrast: perceptual and productive voice onset time characteristics of adults. JASA. 56(3).981-994.

_____, and Roy A. Koenigsknecht. (1976). Development of the voicing contrast: a comparison of voice onset time in stop perception and production. JSHR. 19.93-111.

END

DEPT. OF HEW

NAT'L INSTITUTE OF EDUCATIO

ERIC

DATE FILMED

JULY 31, 197