1242

Distributional Characteristics of VOT in Children's Voiceless Aspirated Stops and Interpretation of Developmental Trends

RESEARCH NOTE

Laura L. Koenig*
Haskins Laboratories
New Haven, CT

Numerous researchers have measured voice onset time (VOT) in children. Authors attempting to trace developmental trends in consonant voicing have frequently framed their hypotheses in terms of how children's VOT means and/or standard deviations compare to adult norms. However, data from previous studies suggest that children's VOTs may not be normally distributed. Specifically, rightward skew is observed in the voiceless aspirated stops, such that mean values exceed the medians. The current work presents detailed distributional analyses of VOTs in /p, t/ from 7 five-year-old children and 14 adults. Distributional non-normality was common in both the adult and child data, as measured by Shapiro and Wilk's *W* statistic. The children showed an insignificant but consistent tendency towards higher values of skew than the adults and greater differences between VOT mean and median values. The results suggest that theories of VOT development should not be based solely on means and standard deviations, but need to address the distributional characteristics of the data more fully.

KEY WORDS: voice onset time, distributional statistics, speech production development

Over the last three decades, researchers have increasingly used acoustic, aerodynamic, and kinematic data to define subtle differences between child and adult speech. To study the development of consonantal voicing, many authors have measured voice onset time (VOT; Lisker & Abramson, 1964) in children, typically relying on means and standard deviations to describe age differences. Where VOT histograms are presented, one can frequently observe that children's productions of voiceless aspirated stops tend to cluster at short positive values of VOT, with a few very long-lag outliers. No published work has explicitly evaluated the extent to which children's VOT data are normally distributed, however. This note reviews evidence for skewed distributions in previous work on children's VOT and presents new data and analyses supporting the conclusion that researchers should take distributional shape into account when reporting and interpreting VOT data from children.

* Currently also affiliated with New York University

1092-4388/01/4405-1058

VOT in Adults and Children

VOT, defined as the time interval between plosive release and the onset of voicing, is one of the primary acoustic cues differentiating contrastive, syllable-initial stop consonants across languages, both in production and in perception (e.g., Abramson, 1977; Abramson & Lisker, 1967, 1973; Lisker & Abramson, 1964, 1970; Zlatin, 1974). Voiceless aspirated stops are measured with positive VOTs, or long voicing lags, because they have a time delay between stop release and voicing onset. Voiceless unaspirated stops show little or no time delay between release and voicing and therefore have VOTs near zero. Fully voiced stops with laryngeal vibration continuing through closure up to the moment of release are measured with negative VOTs. In American English, /p, t, k/ initiating stressed syllables have long positive values of VOT, and /b, d, g/ in the same environment range from negative to short-lag VOTs. Within an individual speaker, VOT values for voiceless aspirated stops vary directly as a function of when peak glottal abduction occurs relative to plosive release (Lisker, Abramson, Cooper, & Schvey, 1969; Löfqvist, 1992; Sawashima, Abramson, Cooper, & Lisker, 1970). Insofar as VOT measures the relative timing between laryngeal and supralaryngeal events, the sequence by which children attain adult-like values of VOT has been widely used as an indication of interarticulator timing control development.

An extensive literature exists on age-correlated changes in VOT, especially for English-learning children. Preston and Yeni-Komshian (1967) and Preston, Yeni-Komshian, Stark, and Port (1968) presented data from naturalistic speech of American-English children beginning at 6 months old. In the early recording sessions, most productions had short-lag VOTs; long-lag stops began to occur more frequently at about age 2. Kewley-Port and Preston (1974) subsequently gathered longitudinal data from American children and found that mean VOTs for /t/ remained short, compared with those of adults, as late as 4.5 years. Inspection of Kewley-Port and Preston's histograms shows that, although most of the children's /t/s fell in the short-lag range, each child produced a few outliers with very long VOTs (up to 290 ms). Zlatin and Koenigsknecht (1976) similarly reported short means for /p, t, k/ in American 2- and 6-year-olds compared with adults. Although the Zlatin and Koenigsknecht article did not discuss distributional characteristics of VOT at length, full distributional statistics for all subjects are available in an appendix of Zlatin's unpublished (1972) thesis. The 1972 statistics show that /p, t, k/ VOT distributions for all subjects tended to have rightward skew, but the skew values were more extreme for the children. The effect of distributional non-normality may be to inflate the p value slightly for a chosen α level (Keppel, 1982), and Zlatin (1972) argued for adopting conservative cutoff points when interpreting statistical results from such data.

Macken and Barton (1980a) used results from children up to 2 years old to formulate a three-step developmental sequence for English-learning children. In Stage I, children showed unimodal VOT distributions in the short-lag range. Secondly, bimodal distributions developed, but averages for the voiceless aspirates remained short relative to those of adults. Stage III was marked by adult-like mean values for long-lag stops. Some children appeared to reach mature values by first "overshooting" (producing very long positive VOT means) and then shortening back to adult values. Barton and Macken (1980) found longer VOT means in 4-year-olds than adults and took this as further evidence for an overshoot phase in acquiring voiceless aspirates. Other authors as well have reported long means and occasional very long (>100 ms) positive values of VOT in children's productions of /p, t, k/ (Gilbert, 1977; Menyuk & Klatt, 1975; Smith, 1978). Studies of older children have yielded no consensus on trends in mean values, but generally agree that high variability may persist in the voiceless aspirated category until age 12 or later (Eguchi & Hirsh, 1969; Kent & Forner, 1980; Ohde, 1985; Ostry, Feltham, & Munhall, 1984).

Although some authors have presented VOT histograms for individual subjects (e.g., Bond & Wilson, 1980; Eilers, Oller, & Benito-Garcia, 1984; Gandour, Petty, Dardarananda, Dechongkit, & Mukngoen, 1986; Kewley-Port & Preston, 1974; Macken & Barton, 1980a, 1980b), other reports of children's VOT have reduced the data to a few summary statistics, most commonly, the mean, range, standard deviation, and (occasionally) the coefficient of variation (Barton & Macken, 1980; Engstrand & Williams, 1996; Kent & Forner, 1980; Menyuk & Klatt, 1975; Ohde, 1985; Smith, 1978; Tyler & Saxman, 1991). Quantitative assessment of VOT category contrast has usually relied on comparison of mean differences (e.g., Barton & Macken, 1980; Macken & Barton, 1980b; Zlatin, 1972), and hypotheses about the development of VOT patterns have frequently made reference to how children's means, standard deviations, and/or ranges compare with adult norms (e.g., Eguchi & Hirsh, 1969; Gandour et al., 1986; Kent, 1976; Kent & Forner, 1980; Macken & Barton, 1980a, 1980b; Smith, 1978; Tyler & Saxman, 1991). If, however, VOT samples are consistently or significantly skewed, then standard parametric statistics are not appropriate, and means and standard deviations do not adequately and accurately represent the data. In particular, because mean values are known to be sensitive to outliers in a distribution (see, e.g., Hays, 1988), greater rightward skew in children's voiceless aspirated VOTs may inflate their means relative to the bulk of productions, overestimate the category differentiation between voiced and voiceless stops, and fail to provide a clear picture of voicing contrast acquisition. The current note presents data on the prevalence and degree of skew in voiceless aspirated stop VOTs recently collected from American English-speaking 5-year-olds and adults. Data from Zlatin's (1972) thesis are discussed for comparison and further support.

Method

Subjects

The data reported here were collected as part of a larger aerodynamic study on voicing in stop consonants and /h/ in men, women, and children, so some methodological details reflect the aims of the overall study (see Koenig, 1998, 2000). Data were collected from 7 speakers in each of three groups: Men (subjects A1–A7, mean age = 35.4 years, range = 26.6–57.7 years), women (A8–A14, mean age = 37.5 years, range = 27.1–51.3 years), and 5-year-old children (mean age = 5.4 years, range = 4.7–5.9 years). The child group consisted of 4 boys (C1–C4) and 3 girls (C5–C7). Given the results of Kewley-Port and Preston (1974) and Zlatin (1972), we expected that 5-year-olds would demonstrate greater VOT variability than adults; our own pilot work (Koenig & McGowan, 1994) further suggested that 5-year-olds could successfully perform the experimental task.

All subjects were normal, healthy, native speakers of American English with no known or suspected history of speech, language, or hearing disorders. Parents of the children filled out questionnaires about the child's developmental history and were questioned to ascertain that there was no history of developmental disorders and no extensive exposure to languages other than English. To verify that the children's language and speech were age-appropriate, a spontaneous speech sample was recorded for each child. Analyses of these data using Miller and Chapman's (1993) Systematic Analysis of Language Transcripts (SALT) programs showed that all children produced mean lengths of utterance (MLUs) within or above expected age-ranges according to Miller and Chapman (1981). A certified speech-language pathologist subsequently reviewed the tapes and verified that the children showed no signs of articulatory or phonological disorders.

Speech Materials

Subjects produced multiple repetitions of 4-syllable utterances of the form [ˌmɑmə ˈCɑpə], in which the target consonant (C) initiated the third, primary-stressed syllable and was one of the set /p, t, b, d, h/. Only the stop data are presented here; data from the full set of utterances are available in Koenig (2000). The experimental utterance was chosen to place the target consonant in a running speech context that 5-year-olds could easily learn and remember. The experimenter presented the utterances to subjects verbally, one at a time, beginning with [ˌmɑmə ˈpʰɑpə], the one most easily interpreted as an ordinary English utterance. The subject produced multiple repetitions of this at a self-selected rate for one or more input trials until many tokens had been collected, and the next utterance was then introduced. No subject showed any signs of difficulty with the task or the novel utterances. The children's attention spans typically constrained the number of productions we could obtain to about 25 tokens each of /p, t/ per subject. Because the larger study was also investigating male-female differences in adults, more tokens were collected from adult subjects to increase the likelihood that subtle sex differences would be evident. To verify that all children had a productive VOT contrast, a small sample of utterances containing /b, d/ was also collected. Finally, because the full aerodynamic study included pressure measurements, more tokens were collected of utterances containing bilabial stops, because these allowed intraoral pressure recording. Table 1 gives the average number of tokens produced per consonant in each age group and the range of tokens across subjects.

Recording Procedures and Signal Processing

For every utterance, oral-nasal airflow was recorded using a Rothenberg mask appropriately sized for the subject. Subjects were instructed to press the mask firmly against their faces during recording so as to prevent air leaks, and a research assistant sat close to the subjects and monitored them visually to make sure that a tight seal was maintained. An acoustic signal was

Table 1. Number (*N*) of tokens produced per subject group per consonant.

Cons.		Adults	Children
/p/	Group total	797	216
	Average *N* per subject	57	31
	Range of *N* across subjects	36–83	18–55
/t/	Group total	443	156
	Average *N* per subject	32	22
	Range of *N* across subjects	10–48	20–25
/b/	Group total	351	76
	Average *N* per subject	25	11
	Range of *N* across subjects	12–59	5–13
/d/	Group total	226	74
	Average *N* per subject	16	11
	Range of *N* across subjects	9–28	7–15

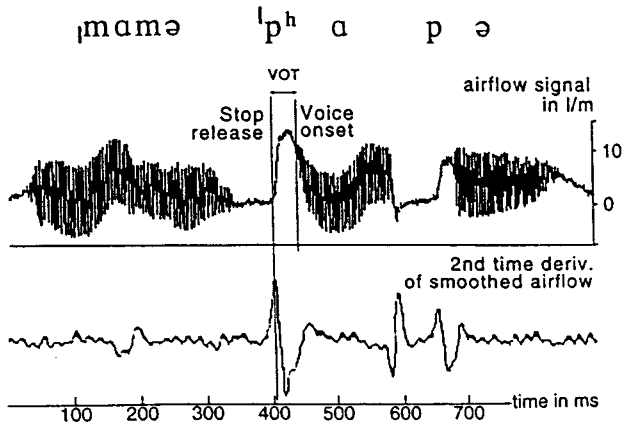
collected using a standing microphone positioned near the subject. Because the acoustic signal received outside the airflow mask is rather damped, all measurements were made from the airflow signals, and the acoustic signal was used only for auditory verification of the utterances. Flow signals were calibrated using a rotameter at the end of each recording session. Acoustic signals were filtered at 9.5 kHz and digitized at 20 kHz, and aerodynamic signals were filtered at 4.5 kHz and digitized at 10 kHz—all with 12-bit quantization.

Flow signals were lightly smoothed two consecutive times with a narrow (5-point) triangular window to eliminate noise and allow easier identification of glottal pulses. A typical signal is shown in Figure 1. A heavily smoothed flow signal was also derived, using a wide (133-point) triangular window to obliterate all or most evidence of glottal pulses. Lastly, the first and second time derivatives of the heavily smoothed flow signals were obtained, using a 3-point difference algorithm, and smoothed with a 133-point window. The first derivative was used to determine the intraoral pressure peaks for /p/ and airflow peaks for /h/ (data reported in Koenig, 2000). The second derivative was used to define oral stop releases, as described below.

Measurement Procedures and Statistical Analysis

The two events that define VOT were measured from the oral airflow signals as shown in Figure 1. Stop releases were set at the local peak in the second derivative of the smoothed flow signal. This represents the rapid increase in airflow that occurs when vocal tract closure is released and corresponds to the plosive burst

Figure 1. Labels used to define stop VOTs. Stop releases were set at peaks in the second time derivative of the heavily smoothed airflow signal. Voice onset was determined visually from the lightly smoothed flow signal, and VOT was calculated as the difference between these two times.



in an acoustic signal. Voicing onset was set at the first visible pulse in an expanded display of the original (lightly smoothed) flow signal, and VOT was calculated as the difference between these two times. To assess measurement reliability, the same experimenter remeasured 5% of the data, using identical procedures, several months after the original measures were made. Correlational analyses for VOT yielded values of r = .985, $p < .0001$.

After VOT measures were calculated, VOT category contrast was tested in each subject's /b, p/ and /t, d/ pairs using the Mann-Whitney U statistic, which, unlike the t test, does not assume distributional normality. Summary statistics were also computed for each subject's VOTs. These included the mean, median, standard deviation, coefficient of variation, skew, skew/standard error (skew/SE), and Shapiro and Wilk's W statistic. The skew/SE statistic normalizes for group differences in standard deviation and sample size and corrects for possible artifacts resulting from the fewer number of tokens collected from the children. The W statistic is a general test for distributional normality (see discussion in Dixon, 1992). These summary statistics were entered as the dependent measures into Analyses of Variance (ANOVAs), with group (men, women, and children) as a cross-subject factor and consonant (/p/ or /t/) as a within-subject factor. On the measures discussed here, there were no significant differences between men and women, so the adults were combined into a single group and all analyses were rerun with two groups (adults vs. children). To correct for the fact that ANOVAs were run on 7 dependent variables, the significance level was adjusted according to the Bonferroni principle (Hays, 1988), yielding $\alpha = .007$ (.05/7). Homogeneity of variance analyses showed no significant differences between /p/ and /t/. However, significantly ($p = .01$) greater variances were found for the children than the adults in means, medians, standard deviations, and coefficients of variation. Accordingly, group effects for these variables must be treated with some caution. At the same time, inequality of variances is less of a concern when group sizes are equal (Keppel, 1982), and the original ANOVAs, computed for three groups of equal size (7 each of men, women, and children), showed the same patterns of significance as those reported here.

Results

The U tests confirmed that all subjects had significant ($p \le .001$) VOT differences as a function of voicing category, indicating that the 5-year-olds had all established a productive voicing contrast in initial position. Because /b, d/ were collected merely to verify this, they will not be discussed further.

The ANOVA results are given in Table 2. These show no significant differences between children and adults for means or medians of VOT. Although mean VOTs for /t/ exceeded those for /p/ in most subjects (all but one man, woman, and child), the ANOVA did not yield a significant effect of consonant. Significant group effects were found for standard deviations and coefficients of variation, with the 5-year-olds showing higher values than the adults. (In the original 3-group analyses, Bonferroni/Dunn post hoc tests showed that children differed significantly from women and men on both measures, with $p < .0001$ in all cases). Neither the consonant effect nor the consonant-by-group interaction was significant for either of these measures.

Using a conservative α level, effects for skew and skew/SE did not reach significance, although there was a consistent pattern of higher skew values in the children than in the adults. Individual subject data for skew and skew/SE are given in Table 3. (The *W* statistics in Table 3 are discussed below). Positive skew indicates that most tokens had VOT values toward the short-lag end of the range, with fewer tokens at longer-lag values. This effect can be seen in Figure 2, which shows histograms of /p/ VOT for one adult and one child subject whose skew values differed widely. In the child subject, the VOT data are clustered at fairly short values, with a few long-lag outliers. Table 3 shows that there were no cases of leftward skew (values less than 0) among the child subjects. One effect of positive skew is that the mean will exceed most token values. The magnitude of any such overestimation was assessed by calculating the difference between mean and median values for /p t/ VOTs of each subject. In the adults, no clear pattern emerged: For both /p/ and /t/, the difference was positive for 8 subjects and negative for 6, with values ranging from –2.5 to 3.5 ms and an average difference of .5 ms. For the children, however, VOT means were longer than medians in 6 of the 7 children for each consonant, with values ranging from –1.1 to 8.9 ms and an average difference of 2.7 ms. Although these differences were higher for the children, the ANOVA results failed to reach significance ($F = 7.13, p = .011$).

Table 2. Summary of ANOVA results: distributional statistics for /p t/ VOTs.

Dependent variable	Factor	$F_{(1,38)}$	*p*
Means	Group	.844	.3639
	Cons	2.139	.1518
	Group x Cons	.181	.6729
Medians	Group	.255	.6166
	Cons	2.007	.1647
	Group x Cons	.193	.6626
SDs	Group	62.902	<.0001
	Cons	.025	.8748
	Group x Cons	.666	.4194
CoVs	Group	53.883	<.0001
	Cons	3.948	.0542
	Group x Cons	.178	.6758
Skew	Group	7.064	.0114
	Cons	1.420	.2409
	Group x Cons	1.260	.2686
Skew/*SE*	Group	4.187	.0477
	Cons	2.557	.1181
	Group x Cons	.576	.4526
W statistic	Group	.726	.3996
	Cons	<0.000	>.999
	Group x Cons	7.750	.0083

Note. Group = subject group (adult vs. child); Cons = consonant (/p/ vs. /t/).

Whereas skew varies as the cube of the distance between individual values and the mean, the *W* statistic is based on a covariance matrix between the actual ordered values and a set of ordered values drawn from a normal distribution. Unlike the other summary statistics considered above, each *W* value has an associated *p* value. *W* statistics for all subjects are given in Table 3. Values for *W* may range from 0 to 1, with lower numbers indicating departure from normality and values approaching 1 as the sample becomes more normal. Table 3 shows that non-normality, as measured by significant values of *W*, may be common in VOT measured from running speech, for adult as well as child subjects. The ANOVA showed no significant group effect for the *W* statistic. The group-by-consonant interaction approached significance ($p = .0083$), however, reflecting the fact that values were higher in adults than children for /p/, but higher in children than adults for /t/.

Discussion

Previous studies of VOT in voiceless aspirated stops have agreed that school-age children have higher standard deviations and coefficients of variation than adults, but have disagreed on whether the children's mean values differ from those of adults and on the direction of any effect (Barton & Macken, 1980; Eguchi & Hirsh, 1969; Gilbert, 1977; Kent & Forner, 1980; Kewley-Port & Preston, 1974; Macken & Barton, 1980a, 1980b; Ohde, 1985; Ostry et al., 1984; Tyler & Saxman, 1991; Zlatin & Koenigsknecht, 1976). The current data similarly show greater variability in the 5-year-olds than in the adults, but no differences in mean values.

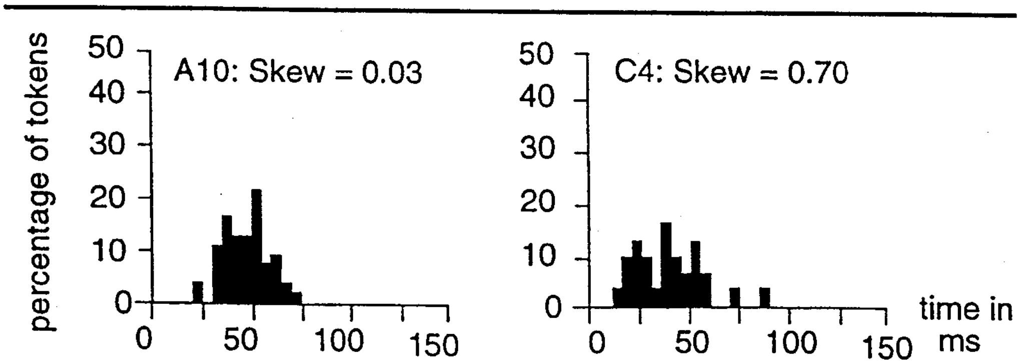
The primary goal of the current analysis was to investigate whether children's VOT values have a consistent pattern of rightward skew. Only one other author

Table 3. Values of Skew, Skew/SE, and the *W* statistic for /p t/ VOTs in all subjects.

	Subjects	Individual data							Group avg.
					Skew				
/p/									
Adults	A1–A7	0.380	0.250	0.390	0.390	-0.050	-0.020	-0.270	0.186
	A8–A14	0.450	-0.360	0.030	0.390	0.430	0.170	0.420	
5-year-olds	C1–C7	0.280	0.950	1.110	0.700	0.460	0.360	0.740	0.657
/t/									
Adults	A1–A7	0.620	-0.310	0.910	-0.030	-0.510	0.420	-0.700	0.184
	A8–A14	0.200	0.520	0.070	0.450	-0.220	0.620	0.530	
5-year-olds	C1–C7	0.030	0.330	0.420	1.160	0.600	0.030	0.010	0.369
					Skew/SE				
/p/									
Adults	A1–A7	1.120	0.864	1.235	1.226	-0.184	-0.050	-0.699	0.602
	A8–A14	1.091	-0.887	0.078	1.353	1.237	0.537	1.510	
5-year-olds	C1–C7	0.846	2.338	2.026	1.559	0.976	0.624	1.662	1.433
/t/									
Adults	A1–A7	0.920	-0.874	2.221	-0.063	-1.447	1.073	-1.627	0.353
	A8–A14	0.396	1.045	0.178	1.106	-0.290	1.155	1.155	
5-year-olds	C1–C7	0.059	0.640	0.803	2.369	1.196	0.048	0.028	0.725
					***W* statistic**				
/p/									
Adults	A1–7	0.960	0.965	0.965	0.933*	0.959*	0.929*	0.938*	
	A8–14	0.941	0.954	0.965	0.950*	0.960	0.973	0.958*	
5-year-olds	C1–C7	0.963	0.899*	0.905*	0.934	0.929	0.862*	0.944	
/t/									
Adults	A1–A7	0.900	0.946*	0.872*	0.833*	0.944*	0.954	0.928*	
	A8–A14	0.952	0.912*	0.979	0.931*	0.937	0.942	0.954	
5-year-olds	C1–C7	0.974	0.943	0.946	0.959	0.958	0.943	0.896*	

Note. *SE* = standard error of the mean. An asterisk for the *W* statistic indicates $p < .05$.

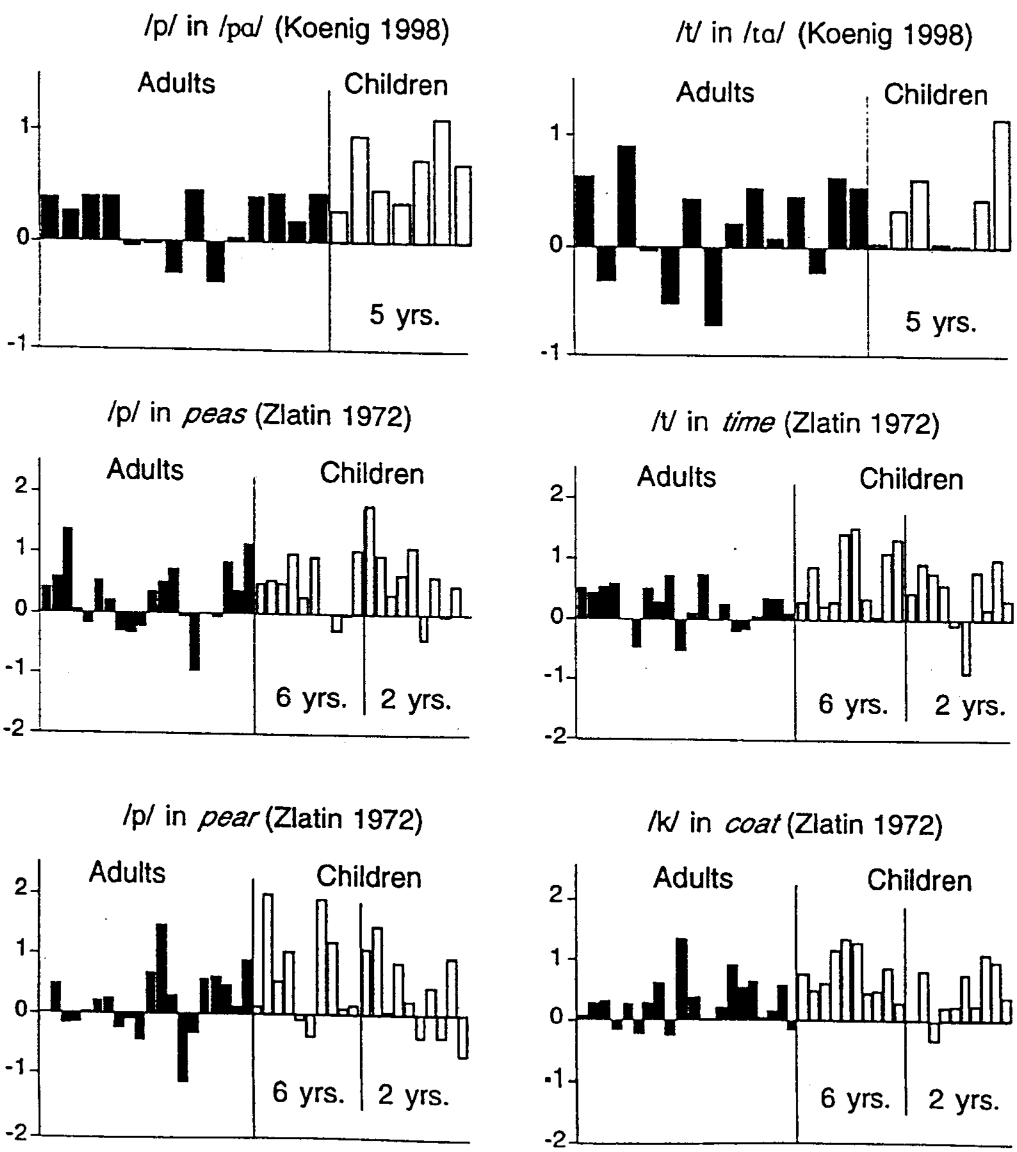
Figure 2. VOT histograms for /p/ in one adult (A10) and one child (C4) subject, showing differences in skew.



has explicitly examined skew measures for VOT (Zlatin, 1972). In Zlatin's study, 10 children at 2 and 6 years and 20 adults were recorded producing 4 minimal pairs differing in initial consonant: *peas-bees*; *pear-bear*; *time-dime*; *coat-goat*. Multiple (30+) repetitions of test words were elicited by means of pictures for the children and printed cards for the adults. This procedure might be expected to yield fairly careful speech, and the mean VOT values obtained for /p, t, k/ were rather long (c. 80–90 ms). For both adults and children, Zlatin (1972) noted a general tendency toward positive skew in voiceless stop VOTs and negative skew in the voiced stops.

Figure 3 compares the voiceless aspirated stop skew values from Zlatin (1972, Appendix H) with those obtained here. Despite the methodological differences between the two studies, the data are consistent in showing that positive skew in /p, t, k/ VOTs is common in adults as well as in children and is generally more extreme in children. An ANOVA on Zlatin's skew values, using the factors group (2-year-old, 6-year-old, and adult) and word (*peas*, *pear*, *time*, and *coat*) yielded a significant group effect (F = 8.443, p = .0003), with post hoc Scheffé tests revealing higher skew values in the 6-year-olds than in the adults (p = .0004). Zlatin's 6-year-olds also showed the highest average skew values of any group in 3 out of 4 words (all but *peas*). Finally, the differences between mean and median values were calculated for Zlatin's data. For the adults, this difference was positive in 66.3% of the cases; for the 6-year-olds and 2-year-olds, it was positive in 77.5% and 80%, respectively.

Figure 3. Skew values for /p t k/ VOTs, comparing the current results with those of Zlatin (1972). Positive skew indicates that the data are clustered towards the short-lag end of the distribution.



Average differences were .7 ms for the adults (range = –4.3 to 7 ms); 2 ms for the 6-year-olds (range = –5.3 to 8.3 ms); and 3.3 ms for the 2-year-olds (range = –4.5 to 12 ms). Thus, means tended to exceed medians in all subject groups, but positive differences were more frequent and extreme in the children than the adults. An ANOVA on these mean-median differences also yielded a significant group effect ($F = 11.781, p < .0001$). Post hoc Scheffé tests yielded $p = .05$ for 6-year-olds versus adults and $p < .001$ for 2-year-olds versus adults. The difference between 2-year-olds and 6-year-olds was not significant ($p = .15$).

In the current data, skew was consistently greater in 5-year-olds than in adults, but the statistical results did not reach significance. Zlatin's data showed significantly higher skew in 6-year-olds than in adults, whereas 2-year-olds did not differ from adults. Taken together, these results suggest that non-normality of VOT distributions may peak at some point during childhood and may reflect a stage of phonetic or phonological acquisition. Because children between 0 and 2 years produce few long-lag values of VOT, regardless of the target category (Bond & Wilson, 1980; Kewley-Port & Preston, 1974; Macken & Barton, 1980a, 1980b; Preston & Yeni-Komshian, 1967; Preston et al., 1968; Tyler & Edwards, 1993; Tyler & Saxman, 1991; Zlatin & Koenigsknecht, 1976), establishing a VOT category contrast in English primarily involves the emergence of tokens with longer-lag values. Higher rightward skew in 5- and 6-year-olds could thus reflect a process by which older children refine their VOT contrasts in the direction of greater differentiation between categories. Long-lag VOT outliers could also arise in children in response to features of child-directed speech. Specifically, some authors have suggested that child-directed speech may exaggerate some phonetic contrasts, including VOT and other durational measures (Bernstein Ratner, 1986; Bernstein Ratner & Luberoff, 1984; Malsheen, 1980; see, however, Baran, Laufer, & Daniloff, 1977; Gurman Bard & Anderson, 1983; and Shockey & Bond, 1980 for counter examples). There is also some evidence that children's productions are sensitive to such features of adult input. Eilers et al. (1984) reported that 2-year-olds imitating adult models produced their longest VOT lags in response to modeled productions with mean VOTs greater than 100 ms. Along similar lines, Tyler and Saxman (1991) measured VOTs in phonologically disordered 2-year-olds undergoing treatment for voicing contrast acquisition and attributed children's very long VOTs for /p, t, k/ (over 300 ms) to the clinician's exaggerated models.

Many previous researchers have argued that measures of variability may be useful in tracing the development of adult-like speech production (e.g., Chermak & Schneiderman, 1986; Eguchi & Hirsh, 1969; Harris, Rubin-Spitz, & McGarr, 1985; Kent, 1976; Kent & Forner, 1980; Leonard, Rowan, Morris, & Fey, 1980; Ohde, 1985; Sharkey & Folkins, 1985; Smith, 1978, 1995; Smith & McLean-Muse, 1986; Smith, Sugerman, & Long, 1983; Tingley & Allen, 1975; Tyler & Saxman, 1991; Zlatin & Koenigsknecht, 1976; but see also Stathopoulos, 1995). The present results suggest that distributional shape may be another important parameter to consider in defining and accounting for developmental trends. The benefit of skew measures is precisely that they are directional and can reveal that variability is most extreme on one side of the distribution.

Although most discussions of developmental trends for VOT have been framed in terms of mean values, the mean may not be the most appropriate measure of central tendency for a quantity that is highly variable and skewed. The lack of consensus among authors about how children's mean VOT values (after age 2) for /p, t, k/ compare with adults' is itself a suggestion that the mean is not a very revealing measure and, as such, is not a good foundation on which to construct detailed theories of voicing acquisition. Median values, which are less sensitive to outliers than means (cf. Hays, 1988), should present a more conservative picture of adult-child differences. More generally, central tendency measures alone may not be the ideal way to evaluate the development of VOT category contrasts. One possible alternative is a measure of overlap between contrastive VOT categories. Zlatin and Koenigsknecht (1976) and Gandour et al. (1986) have both reported decreasing overlap as a function of age.

Rightward skew actually appears to be typical of many durational measures in speech. Kent and Forner (1980) noted that several durational measures taken from sentence productions of children and adults were skewed to the right. (They did not, however, indicate whether this was true for their VOT data.) Crystal and House (1982, 1988) reported on a variety of durations in read speech of English-speaking adults. Their data revealed a pattern of rightward skew for several types of speech sounds and for both fast and slow talkers. Several phone durations in Pols, Wang, and ten Bosch (1996) also show rightward skew. The caveat to be drawn from this is that researchers who reduce durational measures to means and standard deviations run the risk of misrepresenting their data, in that a few extreme outliers may elevate mean values, and standard deviations do not reflect the directionality of the distribution.

Conclusions

The data reviewed here indicate that VOTs for voiceless aspirated stops in English show a consistent pattern of rightward skew. The effect of this is to yield mean values that regularly exceed the value of most productions (that is, the median). Data from Zlatin (1972) show similar trends, despite considerable differences in methodology: Whereas Zlatin collected single-word, citation form productions with varying vowel contexts, the current study placed the consonants of interest in a more fluent running-speech context, with a single flanking vowel. The agreement between the two studies strengthens the conclusion that distributional normality cannot be assumed for VOT data, either for children or adults, though skew may be more extreme in children. Although the elevation of mean values relative to medians is not of great magnitude on average, it can be on the order of 10 ms in some individuals.

The data also suggest the possibility that nonnormality of VOT data may peak at some point during childhood. This warrants further investigation in light of Macken and Barton's (1980a) hypothesized overshoot phase in the acquisition of long-lag stops. Inspection of Macken and Barton's (1980a, 1980b) histograms suggests that some cases in which children had longer means than adults may have reflected occasional long-lag outliers, rather than an overall shift in the center of the distributions. If it is generally true that elevated means in children's speech compared to that of adults result from a few tokens with very long values, then the developmental progression is best characterized as reducing the occurrence of extreme productions and achieving greater consistency of VOTs beyond the short-lag region. The results reported here call for more detailed study into distributional characteristics of VOT data and suggest that additional measures, including skew, may provide useful windows into children's gradual refinement of production strategies for voiceless aspirated stops.

Acknowledgments

The work reported here was supported by NIH grants DC-00865 and DC-00121 to Haskins Laboratories. Portions of these data were presented in an unpublished doctoral thesis submitted to the Department of Speech and Hearing at the CUNY Graduate School and University Center. Special thanks to Marsha Zlatin Laufer for her comments on an earlier version of this article. Thanks also to Alice Faber and Melanie Campbell for assistance in recording the child subjects; to Sylvia Yudicé Walters for reviewing the children's speech samples; to Einar Mencl and Len Katz for advice on statistical treatments; to my subjects and the parents of my child subjects; and to Arthur S. Abramson, Katherine S. Harris, Leigh Lisker, Anders Löfqvist, Christine Sapienza, Richard G. Schwartz, and three anonymous reviewers for their comments on earlier versions of this manuscript.

References

Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica, 34,* 295–303.

Abramson, A. S., & Lisker, L. (1967, Aug.-Sept.). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569–573). Prague: Academia Publishing House, Czechoslovak Academy of Sciences.

Abramson, A. S., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1,* 1–8.

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics, 5,* 339–350.

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

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics, 14,* 303–309.

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics, 12,* 285–289.

Bond, Z. S., & Wilson, H. F. (1980). Acquisition of the voicing contrast by language-delayed and normal-speaking children. *Journal of Speech and Hearing Research, 23,* 152–161.

Chermak, G. D., & Schneiderman, C. R. (1986). Speech timing variability of children and adults. *Journal of Phonetics, 13,* 477–480.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected-speech signals: Preliminary results. *Journal of the Acoustical Society of America, 72,* 705–716.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America, 83,* 1553–1573.

Dixon, W. J. (1992). *BMDP Statistical Software Manual* [to accompany BMDP Release 7]. Los Angeles: University of California Press.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Otolaryngologica, Supplement 57.*

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

Engstrand, O., & Williams, K. (1996, May). VOT in stop inventories and in young children's vocalizations: Preliminary analyses. *Proceedings of Fonetik 96, Swedish Phonetics Conference,* Nässlingen.

Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukngoen, S. (1986). The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language, 13,* 561–572.

Gilbert, J. H. V. (1977). A voice onset time analysis of apical stop production in 3-year-olds. *Journal of Child Language, 4,* 103–113.

Gurman Bard, E., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language, 10,* 265–292.

Harris, K. S., Rubin-Spitz, J., & McGarr, N. S. (1985, Oct.). The role of production variability in normal and deviant developing speech. *The conference on the planning and production of speech in normal and hearing-impaired individuals: A seminar in honor of S. Richard Silverman* (ASHA Reports 15, pp. 50–57). Rockville, MD: American Speech-Language-Hearing Association.

Hays, W. L. (1988). *Statistics* (4th ed.). Orlando: Holt, Rinehart & Winston.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research, 19,* 421–447.

Kent, R. D., & Forner, L. L. (1980). Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics, 8,* 157–168.

Keppel, G. (1982). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Kewley-Port, D., & Preston, M. (1974). Early apical stop production: A voice onset time analysis. *Journal of Phonetics, 2,* 195–210.

Koenig, L. L. (1998). *An aerodynamic study of consonantal voicing control in normal men, women, and 5-year-olds.* Unpublished doctoral dissertation, CUNY Graduate School and University Center.

Koenig, L. L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research, 43,* 1211–1228.

Koenig, L. L., & McGowan, R. S. (1994). *A case study into the aerodynamics of consonant production in running speech of children.* Poster presented at the 127th meeting of the Acoustical Society of America, Cambridge, MA.

Leonard, L. B., Rowan, L. E., Morris, B., & Fey, M. E. (1980). Intra-word phonological variability in young children. *Journal of Child Language, 9,* 55–69.

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

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

Lisker, L., Abramson, A. S., Cooper, F. S., & Schvey, M. H. (1969). Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America, 45,* 1544–1546.

Löfqvist, A. (1992). Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants. *Language and Speech, 35,* 15–28.

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

Macken, M. A., & Barton, D. (1980b). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language, 7,* 433–458.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology. Volume 2: Perception* (pp. 173–184). New York: Academic Press.

Menyuk, P., & Klatt, M. (1975). Voice onset time in consonant cluster production by children and adults. *Journal of Child Language, 2,* 223–231.

Miller, J. F., & Chapman, R. S. (1981). Research note: The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24,* 154–161.

Miller, J. F., & Chapman, R. S. (1993). *SALT: Systematic analysis of language transcripts.* Madison, WI: Waisman Center.

Ohde, R. N. (1985). Fundamental frequency correlates of stop consonant voicing and vowel quality in the speech of preadolescent children. *Journal of the Acoustical Society of America, 78,* 1554–1561.

Ostry, D. J., Feltham, R. F., & Munhall, K. G. (1984). Characteristics of speech motor development in children. *Developmental Psychology, 20,* 859–871.

Pols, L. C. W., Wang, X., & ten Bosch, L. F. M. (1996). Modelling of phone duration (using the TIMIT database) and its potential benefit for ASR. *Speech Communication, 19,* 161–176.

Preston, M., & Yeni-Komshian, G. (1967). Studies on the development of stop-consonants in children. *Haskins Laboratories Status Report on Speech Research, SR11,* 49–52.

Preston, M. S., Yeni-Komshian, G., Stark, R. E., & Port, D. K. (1968). Developmental studies of voicing in stops. *Haskins Laboratories Status Report on Speech Research, SR13/14,* 181–184.

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

Sharkey, S. G., & Folkins, J. W. (1985). Variability in lip and jaw movements in children and adults: Implications for the development for speech motor control. *Journal of Speech and Hearing Research, 28,* 8–15.

Shockey, L., & Bond, Z. S. (1980). Phonological processes in speech addressed to children. *Phonetica, 37,* 267–274.

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics, 6,* 37–67.

Smith, B. L. (1995). Variability of lip and jaw movements in the speech of children and adults. *Phonetica, 52,* 307–316.

Smith, B. L., & McLean-Muse, A. (1986). Articulatory movement characteristics of labial consonant productions by children and adults. *Journal of the Acoustical Society of America, 80,* 1321–1328.

Smith, B. L., Sugerman, M. D., & Long, S. H. (1983). Experimental manipulation of speaking rate for studying temporal variability in children's speech. *Journal of the Acoustical Society of America, 74,* 744–749.

Stathopoulos, E. T. (1995). Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech. *Journal of Phonetics, 23,* 67–80.

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development, 46,* 186–194.

Tyler, A. A., & Edwards, M. L. (1993). Lexical acquisition and acquisition of initial voiceless stops. *Journal of Child Language, 20,* 253–273.

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

Zlatin, M. (1972). *Development of the voicing contrast: A psychoacoustic study of voice onset time.* Unpublished doctoral dissertation, Northwestern University.

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America, 56,* 981–995.

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research, 19,* 93–111.

Received November 11, 2000

Accepted June 5, 2001

DOI: 10.1044/1092-4388(2001/084)

Contact author: Laura L. Koenig, PhD, Haskins Laboratories, 270 Crown Street, New Haven, CT, 06511.
E-mail: koenig@haskins.yale.edu