Published in final edited form as:

J Phon. 2007 July 1; 35(3): 341–352. doi:10.1016/j.wocn.2006.10.001.

VOT in the babbling of French- and English-learning infants

D. H. Whalen,

Haskins Laboratories, 300 George St., New Haven, CT 06511

Andrea G. Levitt, and

Haskins Laboratories, 300 George St., New Haven, CT 06511, and Dept. of French, Wellesley College, Wellesley, MA 02481

Louis M. Goldstein

Haskins Laboratories, 300 George St., New Haven, CT 06511, and Dept. of Linguistics, Yale University, New Haven, CT 06520

Abstract

Different languages use voice onset time (VOT) in different ways to signal the voicing contrast, for example, short lag/long lag (English) vs. prevoiced/short lag (French). Also, VOT depends on place of articulation, with labial VOTs being shorter than velar and alveolar and, sometimes, alveolar being shorter than velar. Here we examine the VOT in babbled utterances of five French-learning and five English-learning infants at ages 9 and 12 months. There was little or no difference between the languages for duration of positive VOTs, which were usually in the “short lag” range. The duration of prevoicing also did not differ between languages, but the proportion of prevoiced utterances did (French-learning infants: 44.2% prevoicing; English-learning: 14.3%). Labial, alveolar and velar stops differed in VOT, with alveolar longer than labial and velar longer than alveolar, suggesting a mechanical cause. The lack of long-lag VOT indicates that the English-learning infants have not mastered aspiration by 12 months. The different proportions of prevoicing, however, suggest that the French-learning infants attempt to imitate the prevoicing that is used frequently (and contrastively) in their native language environment. The results suggest that infants are sensitive to the voicing categories of the ambient language but that they may be able to control prevoicing more successfully than aspiration.

1. Introduction

The distinctive use of voicing for stops in initial position is one of the most commonly used contrasts across languages (approximately 2/3 of the 317 languages in Maddieson (1984)). In initial position, the most fruitful way of describing the difference acoustically has been in terms of voice onset time, or VOT (Abramson & Lisker, 1970; Lisker & Abramson, 1964, 1971), the time between the onset of laryngeal vibration and the release of a stop. When the stop is released prior to voice onset, the VOT is considered to be positive (as with aspirated stops), while voicing onset that precedes the release is considered negative VOT. For languages that have only one contrast that is called “voicing,” one of two patterns is typically used: either a prevoiced/short lag difference (e.g., French, Spanish) or a short lag/long lag difference (e.g., English, German). As a robust and common distinction, voicing provides a convenient tool for examining the attunement of children to their language environment: The voicing categories

will be heard frequently, and the cross-language differences allow us to look for evidence of systematic differences even before meaningful words are acquired.

Children acquire competence in producing the voicing distinction well after they begin to produce words. Even though typical early words in English contain voiced (“bottle”, “baby”) and voiceless (“papa,” “please”) stops, systematic control of the distinction in production may not be achieved until the second birthday (Gilbert, 1977; Macken & Barton, 1980a; Snow, 1997; Tyler & Edwards, 1993; Zlatin & Koenigsnecht, 1976), although the beginning of distinct production has been reported around 3 months after the onset of meaningful word production (Stoel-Gamon, 1985). Until the distinction is mastered, the most typical early productions are short-lag VOT, that is, voiceless unaspirated (e.g., Jakobson, 1968; Kewley-Port & Preston, 1974). As with these early English results, measurements from Spanish-learning infants are also clustered at the short-lag value (Eilers, Oller, & Benito-Garcia, 1984). Swiss-German-learning infants show a similar pattern (Enstrom, 1982). A transcriptional analysis of Thai indicated that aspirates occur only sporadically by 10 months, more consistently after that (Tuaycharoen, 1979). Early word imitation by a Cantonese-learning infant (1;7 to 2;3) was found not to differentiate the aspirated and unaspirated categories of that language (Clumeck, Barton, Macken, & Huntington, 1981); instead, values like the adult unaspirated category were used. Thus production of short-lag VOTs precedes that of aspirated and pre-voiced stops. Traces of this original preference are evident in English even at five years of age, when the average VOT values for the voicing contrast correspond to the adult values, but children produce more short-lag utterances than adults (Koenig, 2001). Further, later mastery of prevoicing compared to short-lag VOT occurs in children acquiring such languages as Spanish (Eilers et al., 1984; Macken & Barton, 1980b), French (Allen, 1985) and Thai (Gandour, Petty, Dardarananda, Dechongkit, & Mukngoan, 1986), which all include a distinctive prevoiced/short-lag contrast. (Thai also includes distinctive aspirates.) Current evidence, then, points toward a fairly universal pattern of acquisition of VOT.

The exact VOT characteristics of the input to young language learners are not clear. The VOTs in infant-directed talk (IDT) have been found to differ from those of adult-directed talk (ADT) for Swedish (Sundberg & Lacerda, 1999). IDT addressed to three-month-olds had shorter VOTs than ADT. Since the speech that infants pay the most attention to is IDT (Fernald, 1984), the input might be less different across languages than the adult language would lead us to expect. Cross-language differences, then might be weakened or nonexistent if this pattern found for Swedish turns out to be widespread. On the other hand, some studies have shown that IDT varies with the linguistic level of the child. Bernstein Ratner (1984) found vowel space differences in speech directed to children aged 2 to 3;6. In a finding more related to the present investigation, Malsheen (1980) found that mothers made the voicing contrast in initial consonants more distinct by ensuring that there was no overlap in VOT values for voiced and voiceless stops when speaking to children at the one-word stage, but not before or after. Children are also exposed to a great deal of adult language, and so they may also be sensitive to the patterns of VOT in the language as a whole. Babbling shows a great deal of sensitivity to the ambient language (see Boysson-Bardies, 1999 for a review), and it may be that some language-specific VOT patterns begin to emerge in babbling.

Adult VOT values vary systematically as a function of place of articulation. Labial VOTs are consistently shorter than those for lingual stops. There appears to be some speaker variation as to whether alveolar stops have a shorter VOT than velars (Lisker & Abramson, 1967; Nearey & Rochet, 1994; Weismer, 1979; Zue, 1976) or whether they are the same (Cooper, 1991; Crystal & House, 1988; Docherty, 1989). Most accounts of these differences assume that there is an automatic mechanism at work. For unaspirated stops, the different rates at which the stop closures are released seem to be sufficient to explain the approximately 20 ms difference in VOT (e.g., Cho & Ladefoged, 1999; Maddieson, 1997). The lips separate quickly, both because

of their relatively small mass and the greater movement of the jaw at that point relative to the velar region. Velar contact is slow to release and may show re-closing (giving a “double burst”). Coronal contact is accomplished in several ways (e.g., apical vs. laminal), giving inconsistent results for that place of articulation as compared to velars. If the VOT differences are automatic, then we could expect to find them in babbled stops as well. In fact, labial VOTs were significantly shorter than velar ones in utterances produced by Spanish- and English-learning infants (Eilers et al., 1984), at least when negative VOTs are included in the means. A similar pattern was found for Swiss-German-learning infants (Enstrom, 1982), although no statistical tests were applied. Further, several developmental studies have been done on early lexical production and indicate that this dependence on place of articulation exists from the earliest word productions onward (see ages two and up in Table 1). Although there is variation in the effect of the alveolars, the velars always have longer VOTs than the labials.

Our goal in the present study was to see if there was evidence in the babbling of French- and English-learning infants for early attunement to the voicing characteristics of their target language and to see whether the difference due to place of articulation appears in these earliest stops. Language-specific tuning of laryngeal timing for the voicing contrast would be surprising given the late appearance of within-language contrasts, at least as far as average VOTs are concerned. Eilers et al. (1984) report differences in the proportion of voicing lead in Spanish-learning infants and long lag in English-learning infants, but do not test those for significance. We will test for similar differences in distribution here.

2. Methods

2.1. Participants

The utterances of 6 American and 6 French infants were recorded weekly in the home by their parents. The French-learning infants lived in Paris, and the American English-learning infants resided in the northeastern United States. The parents of all infants were middle class and college educated. Recordings were made on cassette taperecorders using high quality microphones. The recording sessions each lasted between 10–20 minutes. Parents were instructed to choose a time for recording when the infant was alert and unlikely to cry. We attempted to have the babbling be free from imitation. If the child was not babbling spontaneously, parents were asked to elicit babbling by playing with toys, gesturing, etc. If they spoke during these non-babbling intervals, parents were asked to stop speaking once the infant began to vocalize. The microphone was positioned about 20 cm from the infant while recording. Each session was accompanied by a comment sheet indicating the date, time, and situation (e.g. “in bath”) of the recording.

The utterances were digitized via the Haskins Laboratories PCM system, (Whalen, Wiley, Rubin, & Cooper, 1990) with low-pass filtering at 9.8 kHz and a sampling rate of 20 kHz. The utterances from the 6, 9, and 12 month recordings were input (see Whalen, Levitt, & Wang, 1991 for more details on ages for all the recording sessions). For some of the infants, recordings were not possible at each month indicated. JZ and MB, both French infants, were missing recordings. JZ was missing the 9 and 12 month recordings, and MB the 12 month recording. YC, who was also French, did not have 12 month recordings, so her 11 month recordings were substituted. Two of the infants had very few utterances at various months. In this case, recordings from an adjacent month were used to supplement those months. For English subject MA, 25.8% of the 6 month data was from the 6 month recordings, and the remaining 74.2% was taken from the 7 month recordings. For the same subject, 37.7% of the 12 month data was from the 12 month recordings, and the remaining 62.3% was from the 11 month recordings. For subject MB, a French infant, 35.0% of the 6 month data was from the 6 month recordings, while the remaining 65.0% was taken from the 7 month recordings.

The infants' vocalizations were divided into breath groups. A breath group is defined as a sequence of syllables separated from adjacent utterances by at least 750 ms of silence. Those vocalizations that were speech-like (this excludes vegetative noises, squeals, growls and emotive sounds) were transcribed. The utterances of all twelve infants were transcribed into the IPA by two phoneticians, one a native speaker of Mandarin Chinese (the primary transcriber) and the other a speaker of Belgian French (the secondary transcriber); both were also fluent in English. The primary transcriber was also fluent in Russian, with no substantial exposure to French; his specific language background was happenstance. His transcriptions were considered primary only because he had completed the transcription of the database at the beginning of our measurement sessions. The transcriptions yielded a total of 9174 utterances from the 12 infants. There were 962 which had initial stops in the primary transcription. The stops were categorized in relation to both transcriptions, but any stops that might have been labeled by the secondary transcriber but not the first were excluded by our procedure.

One French-learning infant (JZ) did not have utterances at the 9 and 12 month ages studied, so he was dropped from the experiment. In order to balance the cross-language results, an English-learning infant (VB), chosen at random, was dropped as well, leaving five from each language environment. The active vocabulary of the infants was not assessed.

2.2. Stimuli

All utterances from the 9 and 12 month recordings for which the transcription began with a stop were examined. (The 6 month recordings had too few stops for many of the infants to provide reliable data.) The majority of these were measured for VOT (positive or negative). In some cases, the signal was too weak or there was extraneous noise that made measurement impossible. The transcriptions of the primary transcriber were used as the basis for categorizing initial stops according to place of articulation. The two transcriptions agreed on the stop place of articulation 44.7% of the time (53.8% for labials, 44.1% for alveolars, and 36.2% for velars). 24.3% of the cases had a stop of a different place, 19.0% had no consonant in the second transcription, and the remaining 12.0% had a fricative or glide in the secondary transcription. While this level of agreement appears lower than other reported levels (e.g., the 76.8% for the consonants in Davis & MacNeilage, 1995), several factors are at play here. First, we attempted to transcribe every utterance in our sessions, while many studies are limited to selections (e.g., canonical babbling in Davis & MacNeilage, 1995). Second, we did not attempt to reconcile differences (as in, e.g., Blake & Boysson-Bardies, 1992). Third, we kept all tokens while some studies reject items that were not agreed upon without giving any indication of how frequently that happened (e.g., Eilers et al., 1984; Enstrom, 1982). Fourth, our transcribers, though trained in the IPA, had different language backgrounds, which has been shown to lower levels of agreement (Boysson-Bardies & Vihman, 1991).

2.3. Procedure

All initial stops identified in the primary transcription were analyzed, excluding glottal stops. Stops that were labeled as ambiguous in the transcription were placed into the first category of the two written in the transcription, without any indication of the uncertainty. There were no instances of uvular or pharyngeal stops, so all stops were labeled as labial, alveolar or velar. To analyze the VOTs according to the transcription of the secondary transcriber, all utterances that were transcribed with a stop by both transcribers were selected. Note that this means that there were initial stops in the secondary transcription that were not analyzed; only the subset that overlapped with the primary transcriber were analyzed.

VOTs were measured primarily from visual inspection of the waveform aided by careful listening. This method has been found to be the most reliable means of determining VOT from

the acoustics (Francis, Ciocca, & Yu, 2003). Display was accomplished with the Haskins Laboratories HADES program (Rubin, 1995). In some instances, a spectrogram was also examined. Prevoicing was apparent as a nearly sinusoidal signal preceding the formants of the vocalic segment, even in those cases where there was no clear release burst. Aspiration was apparent in the aperiodic portion of a signal before the voicing of the vocalic segment.

2.4. Ancillary measures

A subset of the French subjects' productions were retranscribed and measured by a different set of listeners. A limited set of tokens from speakers EC (10 tokens), MS (33 tokens) and YC (10 tokens) were analyzed. Selected utterances that were transcribed as having stops in initial or medial position were measured. VOT was measured in a way similar to that in the main experiment.

3. Results

The VOT measurements can be seen in Tables 2 and 3, and Figure 1. These represent averages of all the positive VOT values; negative values were excluded. The values were submitted to an ANOVA with the factors Language, Age (9 or 12 months) and Place of Articulation, separately for each transcriber. Individual VOT measurements were treated as cases so that the contribution from each child would be proportional to his or her actual amount of data; thus, all factors were treated as between factors. As can be seen in the figure, there is virtually complete overlap between the two languages. This is reflected in the lack of a main effect of Language ($F(1,726) < 1$, n.s.) for the primary transcriber. However, for the secondary transcriber, the main effect of Language was significant ($F(1,495) = 5.26$, $p < .05$), with the French-learning infants having values about 8 ms shorter on average.

Age is significant as a main effect ($F(1,726) = 5.14$, $p < .05$), though it was marginal for the secondary transcriber ($F(1,495) = 3.84$, $p = .0506$). The interaction with Language is not significant ($F(1,726) = 1.56$, n.s.; $F(1,495) < 1$). The VOTs in general were somewhat shorter at 9 months of age than at 12 (see Table 4). Similar results are found in the secondary transcriber's categorization.

The pattern of evidence in Tables 2 and 3 for Place of Articulation is significant for both transcribers ($F(2,726) = 11.04$, $p < .001$; $F(2,495) = 3.51$, $p < .05$). As with adult speech, VOTs are longer with more posterior places of articulation. There was no effect of Language on the Place effect ($F(2,726) < 1$, n.s. for the interaction; $F(2,495) = 1.44$, n.s.) nor of Age on Place ($F < 1$ for both transcribers). The three-way interaction also was not significant ($F < 1$ for both transcribers).

Although the means for the alveolar stops are longer than those for the labial, it is not immediately apparent whether they are statistically different. Two analyses were conducted on the results from the primary transcriber, one contrasting labial and alveolar, and another contrasting alveolar and velar. The alveolar/velar difference was significant ($F(1,656) = 25.42$, $p < .001$) while the labial/alveolar difference was only marginally significant ($F(1,436) = 2.81$, $p < .10$).

The relative frequency of prevoicing in the two languages can be seen in Table 5 for the primary transcriber. An ANOVA with the factors Language and Place of Articulation (3 levels) was performed on the percentage of instances which were prevoiced for each of the five infants in each language environment. (Age was dropped as a factor since there were too few instances for a reliable assessment of age.) Language was a significant main effect for both transcribers ($F(1,8) = 25.72$, $p < .001$; $F(1,8) = 73.20$, $p < .001$), confirming the reliability of the sizable difference in rate of prevoicing (14.3% for English, 43.2% for French). Place was not

significant as a main effect (both $F(2,16) < 1$, n.s.) or in an interaction with Language (both $F(2,16) < 1$, n.s.).

The duration of the prevoicing itself was 43.2 ms on average for the English-learning infants and 38.6 for the French-learning ones (see Table 6). There were not enough tokens to do a reliable analysis of further effects (e.g., for place of articulation).

For the ancillary measurements of a limited number of tokens from three infants, positive VOTs averaged 8.5 ms, and the percentage of prevoiced tokens was 54.9%. While not as many measurements were made (and an insufficient number for a statistical test), the general tendencies are still apparent. There were primarily short VOTs, and there was a relatively large percentage of prevoicing for these French babblers.

4. Discussion

The initial stop productions in a babbling data base were examined for language-particular and universal aspects of voicing, and patterns of both types were obtained. The duration of positive VOTs was similar on average for the two languages when categorized by our primary transcriber and similar to those of other studies. When categorized by the secondary transcriber, however, the VOTs were slightly longer (8 ms) for English than for French. There was a slight increase in VOT for both language environments between the ninth and twelfth months, which was significant for the primary transcriber and marginally significant for the secondary. The percentage of utterances with prevoicing, on the other hand, was found to differ between French and English for both transcribers, in ways consistent with the target languages. The results show the continuing interplay of general and specific processes in the babbling stage of language acquisition.

The general tendency to have short-lag VOTs in babbling is fairly strong. It has been found for languages like Spanish that have a prevoiced/short-lag contrast (Eilers et al., 1984), for languages like Swiss German and English that have a short-lag/long-lag contrast with optional prevoicing (Enstrom, 1982 and the present results), and even languages like Cantonese that have a short-lag/long-lag distinction with virtually no prevoicing (Clumeck et al., 1981). Despite the large differences in positive VOT between French and English in adult language (Lisker & Abramson, 1964), there was no difference due to language environment for the main transcriber's results, and less than 10 ms difference for the second transcriber.

The duration of the negative VOTs in our babbling results is not as large as that found in languages that use negative VOTs distinctively or habitually. For example, the average across the nine languages (including English, with its optional prevoicing) of Lisker and Abramson (1964) is -91 ms. Values reported for French are a similar -95.7 ms in one study (Nearey & Rochet, 1994) and a somewhat longer -110 ms in another (Hazan & Boulakia, 1993). These numbers compare with -41 ms for our babblers. It is possible that this indicates that the prevoicing tends to be less adult-like in just the way that the positive VOTs are (given that they are shorter than aspirated categories in the adult language). However, it is equally possible that the stop closures were shorter for the babbling and that the full closure was voiced. Since we have no other acoustic indication of stop closure formation in these utterance-initial stops, we cannot be certain whether the prevoicing shows greater or lesser similarity to the adult implementation. Babbling tends to have differences from adult language (as with the positive VOTs), so it would seem likeliest that there is a somewhat different mechanism at work here. Whatever is responsible, it is being more successfully implemented by the French-learning babblers than the English-learning ones.

Another universal effect, changes in VOT due to place of articulation, was also present in our babbling results. This lends credence to the explanations of this effect, at least for unaspirated

stops, as an automatic consequence of articulation (e.g., Cho & Ladefoged, 1999; Cooper, 1991; Klatt, 1975; Maddieson, 1997). The shorter time that it takes after release of labials for the pressure difference between the subglottal and oral cavities to return (thus allowing voicing to resume) results in a shorter VOT compared to alveolar and velar stops. It is always possible, of course, that the babbling results could be influenced by imitation of the place of articulation pattern as well, since the same pattern does occur in the target languages. However, since the long VOTs of the aspirated stops were not imitated, it seems unlikely that much smaller place of articulation pattern is imitated while the larger voicing distinction is not. Another possible explanation for this pattern of results is that it is due to the transcription rather than the actual articulation. It could be that the transcribers' perception of the place of the stop was influenced by the VOT. Longer VOTs might have made the stops sound more like velars, while shorter ones might have sounded more like alveolars and labials. This does not seem to be the case here, since there was a large degree of overlap in the ranges of VOT for the three places of articulation. That we found no modification of this effect due to language environment may be somewhat surprising given Nearey and Rochet's (1994) assessment that "for voiceless stops, context dependent effects in production appear to be larger for French than for English" (p. 7). However, they were more concerned with the effects of vowel quality on VOT, an aspect for which we did not have enough data to justify an investigation. Further, they studied 5- and 13-year-old children rather than the infants analyzed here. In general, the data support the universality of the place of articulation effect.

It is all the more surprising, then, that a language difference appeared in the percentage of utterances that were prevoiced. It takes many years for the distinction to be adult-like (Allen, 1985; Eilers et al., 1984; Gandour et al., 1986), yet the babbling reflected the proportions in spoken language remarkably well. The overall percentage of prevoicing seen in the English-learning infants' babbling (14.3%) was similar to the values found for English/g/with adult (17.6%) and two-year-old (14.4%) speakers in Davis (1995). (Only/g/was analyzed there due to the need to have solid comparisons with her Hindi results.) Another similarity is found in the results of Enstrom (1982), where a percentage of 6.7% prevoicing can be calculated from the figures. This compares with a near absence of prevoicing in her measurements for adults. The 43.2% prevoicing in the French-learning infants' babbling was remarkably close to what might be expected on the basis of phoneme frequencies. About 44% of the stops in French are voiced (Delattre, 1965), and 91% of adult French productions have prevoicing (Nearey & Rochet, 1994). These two frequencies multiply out to almost exactly what was obtained here. However, this should be interpreted with restraint since the percentage is also similar to that found for two-year-old Hindi learners (Davis, 1995), and we do not have similar frequency counts for that language. Further, they stand in contrast to Allen's (1985) finding that young French speakers often add an unnecessary syllable before a voiced stop to allow the voicing to continue. The French speakers may be more similar to Macken and Barton's (1980b) Spanish learners, who often used spirantized variants instead of true stops in productions that normally contain voiced consonants. In any event, the early attunement that we see in the babbling may undergo a reorganization as lexical distinctions begin to be made.

The characteristics of the prevoiced stops may differ from those of adult productions as well. Phenomenologically, the analyzers of the ancillary measurements found it difficult to assign an instant of release to many of the prevoiced stops. It sounded as though the French infants were not only prevoicing these stops but were also releasing them in a way that made it difficult to find the time at which the articulators forming the closure came apart. Examples are shown in Figures 2 and 3. The shaded regions of both figures exhibit a low rising amplitude profile, which would be typical of prevoicing. Yet the spectrum is clearly not that of prevoicing, which would show energy only at very low frequencies. Here, there is energy at the formant frequencies indicating that the mouth is at least partially open. Note also that the spectrogram exhibits no obvious release burst. This would be consistent with an incomplete oral closure

and/or with a very gradual mouth opening. It may be, then, that the French infants were imitating something that sounds (partially at least) like prevoicing, using a very different mechanism than is employed by adult speakers. Since the prevoicing category is learned relatively late (Allen, 1985; Eilers et al., 1984; Gandour et al., 1986; Macken & Barton, 1980b), this would not be surprising. No matter what the mechanism, the French-learning infants did have significantly more prevoiced utterances.

The stops with a positive VOT, then, could be those stops that achieved a more adult-like closure, which would tend to make continued glottal activity difficult to maintain. The appearance of small positive VOTs in babbling would not necessarily indicate that a devoicing gesture was being coordinated with the stop closure, if we assume that glottal pulsing stopped for aerodynamic reasons. Longer VOTs (in the aspiration range) would indicate the coordination of two gestures (a closure and a devoicing) and seldom occur in our babbled utterances. Results in the literature suggest that various strategies are used in acquiring voicing distinctions and that the achievement of an adult-like pattern takes many years.

It may be that the appearance of truly voiced or truly voiceless stops in babbling is rather rare, and that it is much more common for the voicing to depend on the aerodynamics involved in the stop closure. This would mean that the later implementation of both aspiration and prevoicing would be novel utterance types and not direct continuations of organizations learned in babbling. Both would, in fact, require the infant to control patterns of intergestural (laryngeal-oral) coordination which has been hypothesized to begin three months after the onset of meaningful word production, at the earliest (Studdert-Kennedy & Goldstein, 2003). Ideally, a noninvasive means of measuring infants' articulators could be used to discover whether, for example, the cessation of voicing had the appearance of an adult devoicing gesture or whether it looked more like a dampening of oscillation without any overt control. Such instruments are not available to us now, but the acoustic measurements presented here allow us to speculate that the control of voicing is quite different in babbling than in early language, and that place-related changes in VOT are physiologically- and/or aerodynamically-based.

Acknowledgments

This research was supported by NIH grant DC-00403 to Haskins Laboratories. Additional help with the stimuli was provided by Winifred McGowan, Michele Sancier, Pai-Ling Hsiao, and Iris Smorodinsky. Ancillary measurements were made by members of the Development of Phonology class at Yale University, spring 2000. We thank Catherine T. Best and three anonymous reviewers for helpful comments.

References

- Abramson, AS.; Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In: Hála, B.; Romportl, M.; Janota, P., editors. *Proceedings of the 6th International Congress of Phonetic Sciences*, Prague 1967. Prague: Academia; 1970. p. 569-573.
- Allen GD. How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language* 1985;12:37–46. [PubMed: 3980606]
- Bailey PJ, Haggard MP. Perception-production relations in the voicing contrast for initial stops in 3-year-olds. *Phonetica* 1980;37:377–396. [PubMed: 7280036]
- Barton D, Macken MA. An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children. *Language and Speech* 1980;23:159–169.
- Bernstein Ratner N. Patterns of vowel modification in mother-child speech. *Journal of Child Language* 1984;11:557–578. [PubMed: 6501465]
- Blake J, Boysson-Bardies Bde. Patterns in babbling: A cross-linguistic study. *Journal of Child Language* 1992;19:51–74. [PubMed: 1551933]
- Boysson-Bardies, Bde. *How language comes to children: From birth to two years*. Cambridge, MA: MIT Press; 1999.

- Boysson-Bardies, Bde; Vihman, MM. Adaptation to language: Evidence from babbling and early words in four languages. *Language* 1991;61:297–319.
- Cho T, Ladefoged P. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 1999;27:207–229.
- Clumeck H, Barton D, Macken MA, Huntington DA. The aspiration contrast in cantonese word-initial stops: Data from children and adults. *Journal of Chinese Linguistics* 1981;9:210–225.
- Cooper, A. An articulatory account of aspiration in English. Yale University; 1991. Unpublished Ph.D. dissertation
- Crystal T, House A. Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America* 1988;83:1553–1573.
- Davis BL, MacNeilage PF. The articulatory basis of babbling. *Journal of Speech and Hearing Research* 1995;38:1199–1211. [PubMed: 8747814]
- Davis K. Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language* 1995;22:275–305. [PubMed: 8550724]
- Delattre, PC. Comparing the phonetic features of English, French, German and Spanish. Heidelberg: Julius Groos Verlag; 1965.
- Deuchar M, Clark A. Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics* 1996;24:351–365.
- Docherty, GJ. An experimental phonetic study of the timing of voicing in English obstruents. University of Edinburgh; 1989. Unpublished Ph.D. dissertation
- Eilers RE, Oller DK, Benito-Garcia CR. The acquisition of voicing contrasts in Spanish and English learning infants and children: a longitudinal study. *Journal of Child Language* 1984;11:313–336. [PubMed: 6746779]
- Enstrom DH. Infant labial, apical and velar stop productions: A voice onset time analysis. *Phonetica* 1982;39:47–60. [PubMed: 7089069]
- Fernald, A. The perceptual and affective salience of mothers' speech to infants. In: Feagans, L.; Garvey, C.; Golinkoff, R., editors. *The origins and growth of communication*. Norwood, NJ: Ablex; 1984. p. 5-29.
- Fischer-Jørgensen, E. ptk et bdg français en position intervocalique accentuée. In: Valdman, A., editor. *Papers in linguistics and phonetics to the memory of Pierre Delattre*. The Hague: Mouton; 1972. p. 143-200.
- Francis AL, Ciocca V, Yu JMC. Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America* 2003;113:1025–1032. [PubMed: 12597195]
- Gandour J, Petty SH, Dardarananda R, Dechongkit S, Mukngoen S. The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language* 1986;13:561–572. [PubMed: 3793815]
- Gilbert JHV. A voice onset time analysis of apical stop production in three-year-olds. *Journal of Child Language* 1977;4:103–110.
- Hazan V, Boulakia G. Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 1993;36:17–38.
- Jakobson, R. *Child language, aphasia and phonological universals*. The Hague: Mouton; 1968.
- Kewley-Port D, Preston MS. Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 1974;2:195–210.
- Klatt DH. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 1975;18:686–706. [PubMed: 1207100]
- Koenig LL. Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends. *Journal of Speech, Language, and Hearing Research* 2001;44:1058–1068.
- Lisker L, Abramson AS. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 1964;20:384–422.
- Lisker L, Abramson AS. Some effects of context on voice onset time in English stops. *Language and Speech* 1967;10:1–28. [PubMed: 6044530]
- Lisker L, Abramson AS. Distinctive features and laryngeal control. *Language* 1971;47:767–785.

- Macken MA, Barton D. The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language* 1980a;7:41–74. [PubMed: 7372738]
- Macken MA, Barton D. The acquisition of the voicing contrast in Spanish: a phonetic and phonological study of word-initial stop consonants. *Journal of Child Language* 1980b;7:433–458. [PubMed: 6969264]
- Maddieson, I. *Patterns of sounds*. New York: Cambridge University Press; 1984.
- Maddieson, I. Phonetic universals. In: Hardcastle, WJ.; Laver, J., editors. *The handbook of phonetic sciences*. Oxford: Blackwell; 1997. p. 619–639.
- Malsheen, BJ. Two hypotheses for phonetic clarification in the speech of mothers to children. In: Yeni-Komshian, GH.; Kavanagh, JF.; Ferguson, CA., editors. *Child phonology, Volume 2: Perception*. New York: Academic Press; 1980. p. 173–184.
- Nearey TM, Rochet BL. Effects of place of articulation and vowel context on VOT production and perception in French and English stops. *Journal of the International Phonetic Association* 1994;24:1–19.
- Rubin, PE. HADES: A case study of the development of a signal analysis system. In: Syrdal, A.; Bennett, R.; Greenspan, S., editors. *Applied speech technology*. Boca Raton, FL: CRC Press; 1995. p. 501–520.
- Snow D. Children's acquisition of speech timing in English: A comparative study of voice onset time and final syllable vowel lengthening. *Journal of Child Language* 1997;24:35–56. [PubMed: 9154008]
- Studdert-Kennedy, M.; Goldstein, L. Launching language: The gestural origin of discrete infinity. In: Christiansen, M.; Kirby, S., editors. *Language evolution*. Oxford: Oxford University Press; 2003. p. 235–254.
- Sundberg U, Lacerda F. Voice onset time in speech to infants and adults. *Phonetica* 1999;56:186–199.
- Tuaycharoen, P. An account of speech development of a Thai child: From babbling to speech. In: Thongkum, TL.; Panupong, V.; Kullavanijaya, P.; Tingsabadh, MRK., editors. *Studies in Tai and Mon-Khmer phonetics and phonology: In honour of Eugénie J. A. Henderson*. Bangkok: Chulalongkorn University Press; 1979. p. 261–277.
- Tyler AA, Edwards ML. Lexical acquisition and acquisition of initial voiceless stops. *Journal of Child Language* 1993;20:253–273. [PubMed: 8376469]
- Tyler AA, Saxman JH. Initial voicing contrast acquisition in normal and phonologically disordered children. *Applied Psycholinguistics* 1991;12:453–479.
- Weismer G. Sensitivity of voice onset measures to certain segmental features in speech production. *Journal of Phonetics* 1979;7:194–204.
- Whalen DH, Levitt A, Wang Q. Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* 1991;18:501–516. [PubMed: 1761611]
- Whalen DH, Wiley ER, Rubin PE, Cooper FS. The Haskins Laboratories' pulse code modulation (PCM) system. *Behavior Research Methods, Instruments & Computers* 1990;22:550–559.
- Zlatin MA, Koenigsknecht RA. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research* 1976;19:93–111. [PubMed: 1271805]
- Zue, VW. Acoustic characteristics of stop consonants: A controlled study. (Technical Report 523). Lexington, MA: Lincoln Laboratory, MIT; 1976.

Figure 1.

VOTs at three places of articulation (as judged by the primary transcriber) for French- and English-learning infants at 9 and 12 months of age. Error bars indicate one standard deviation of the individual means from the global means.

Figure 2. Waveform and spectrogram of an utterance produced by French-learning infant YC at age 1;0. The syllable was transcribed as /ne/ by the main transcriber and as /m9bi/ by the secondary one. The shaded region shows low amplitude rise.

Figure 3. Waveform and spectrogram of an utterance produced by French-learning infant YC at age 11;0. The syllable was transcribed as /ve/ by the main transcriber and as /m9bi/ by the secondary one. The shaded region shows low amplitude rise.

Table 1

Dependence of VOT on place of articulation in English in various earlier works.

English:			
Age	k>t>p	k=t>p	k>t=p
1		(Eilers et al., 1984)	
2	(Macken & Barton, 1980a) (Tyler & Saxman, 1991)(?)		(Zlatin & Koenigsnecht, 1976)
3		(Bailey & Haggard, 1980)	
4	(Barton & Macken, 1980)		
5			
6			(Zlatin & Koenigsnecht, 1976)
7			
Adult	(Lisker & Abramson, 1967) (Klatt, 1975) (Zue, 1976) (Weismer, 1979) (Nearey & Rochet, 1994)	(Crystal & House, 1988) (Docherty, 1989) (Cooper, 1991)	
Other Languages:			
1		Spanish (Eilers et al., 1984) Swiss-German (Enstrom, 1982)	
2		Spanish (Macken & Barton, 1980b) Spanish (Deuchar & Clark, 1996)	
3	Cantonese (Clumeck et al., 1981)		Thai (Gandour et al., 1986)
4			Spanish (Macken & Barton, 1980b)
5			Thai (Gandour et al., 1986)
6			
7			Thai (Gandour et al., 1986)
Adult	French (Fischer-Jørgensen, 1972)		Thai (Gandour et al., 1986) French (Nearey & Rochet, 1994)

Note: Clumeck et al (1981) did not test this aspect of their data, so the ranking of the stop places is an estimate. The distinction between t and p is apparent with the aspirates, marginal with the inaspirates.

Table 2

VOTs for the French- and English-learning infants, main experiment, primary transcriber. Only positive VOTs enter into the average. VOT-ms is the average duration (in ms) of the VOTs for that place of articulation. VOT-% is that same value expressed as a percentage of total syllable duration. N is the number of examples of initial stops for that speaker at that place of articulation.

Speaker English:	Labial		Alveolar		Velar	
	VOT-ms	VOT-%	N	VOT-ms	VOT-%	N
AB	11.6	4.5	17	29.1	8.7	78
CR	17.9	8.9	6	15.9	6.2	34
MA	16.7	6.1	4	26.7	8.8	26
MM	31.4	10.8	2	27.9	8.5	39
NG	24.0	7.3	25	26.0	8.1	82
Average	20.3	7.52		25.1	8.06	
French:						
EC	42.1	11.2	4	66.8	19.7	14
MB	7.4	2.7	2	10.1	2.8	7
MS	14.2	6.3	10	17.5	6.4	19
NM	3.9	5.7	1	8.1	2.5	5
YC	12.6	4.4	7	14.6	5.5	58
Average	16.0	6.1		23.4	7.4	
				30.9	9.0	

Table 3

VOTs for the French- and English-learning infants, main experiment, secondary transcriber. Columns are as in Table 2.

Speaker English:	Labial			Alveolar			Velar		
	VOT-ms	VOT-%	N	VOT-ms	VOT-%	N	VOT-ms	VOT-%	N
AB	19.4	6.4	39	18.2	6.1	30	42.6	13.2	100
CR	17.2	8.2	5	15.3	6.8	23	23.5	7.9	5
MA	13.2	5.1	3	21.3	7.4	22	30.3	9.1	13
MM	24.9	10.3	6	26.9	7.3	23	51.9	14.2	5
NG	28.1	9.0	36	26.3	8.4	69	50.5	16.5	28
Average	20.5	7.8		21.6	7.2		39.8	12.2	
French:									
EC	43.72	11.81	2	37.59	8.75	8	58.26	20.48	4
MB	5.36	2.00	1	12.03	5.06	6			
MS	16.35	6.01	20	16.32	5.48	7	7.35	2.87	1
NM				10.38	3.16	3	8.83	1.49	2
YC	10.26	4.41	19	19.88	6.31	32	27.40	10.89	4
Average	18.9	6.1		19.2	5.8		25.5	8.9	

Table 4
Difference in VOTs between ages 9 and 12 months.

Age	Labial	Alveolar	Velar	Average
English:				
9 month	14.2	24.3	36.4	25.0
12 month	21.1	26.8	38.4	28.8
Difference	6.9	2.5	2.0	4.2
French:				
9 month	14.1	15.1	25.3	18.2
12 month	18.5	25.7	50.2	31.5
Difference	4.4	10.6	24.9	13.3

Table 5

Percentage (and number) of negative VOTs of the French- and English-learning infants, main experiment, primary transcriber. (There was only one labial token for NM.)

Speaker	Labial	Alveolar	Velar	Average
English:				
AB	29.2 (7)	8.2 (7)	10.0 (14)	11.7
CR	14.3 (1)	12.8 (5)	6.3 (1)	11.3
MA	33.3 (2)	21.2 (7)	20.1 (6)	22.1
MM	0.0 (0)	0.0 (0)	29.2 (7)	10.3
NG	26.5 (9)	12.8 (12)	15.7 (17)	16.1
Average	18.53	11	16.26	14.3
French:				
EC	76.5 (13)	36.4 (8)	6.7 (1)	40.7
MB	33.3 (1)	46.2 (6)	14.3 (1)	34.8
MS	33.3 (5)	51.3 (20)	66.7 (6)	49.2
NM	(0.0)	54.5 (6)	40.0 (2)	47.1
YC	30.0 (3)	40.8 (40)	63.0 (17)	44.4
Average	30	45.84	38.14	43.24

Table 6
Duration (and N) of negative VOTs of the French- and English-learning infants,
main experiment, primary transcriber. Averages are weighted by number of tokens.

Age	Labial	Alveolar	Velar	Average
English:				
9 months	29.7 (7)	30.1 (8)	49.0 (17)	40.0
12 months	34.5 (12)	39.8 (23)	53.4(28)	44.9
Average	32.7	37.3	51.7	43.2
French:				
9 months	131.9 (3)	40.9 (33)	57.4 (5)	40.7
12 months	29.3 (19)	32.8 (47)	38.6 (22)	34.8
Average	43.3	36.1	42.1	38.6