Journal of Phonetics (1980) 8, 335–341

318

Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children

Lawrence J. Raphael*
Herbert H. Lehman College, The City University of New York Bronx, New York 10468, U.S.A.

Michael F. Dorman*
Arizona State University Tempe. Arizona 85281, U.S.A.
and

Donna Geffner
St Johns University, Queens, New York 11439, U.S.A.

Received 14th December 1978

Abstract:

Seven consonant-vowel-consonant minimal pairs, differing only in the voicing characteristic of the final consonants, were elicited from 20 three- and four-year old native speakers of American English. Spectrographic analyses of the utterances revealed that (1) children produced vowel duration differences of the same nature and magnitude as those found in adult speakers' utterances; (2) the duration of a preceding vowel, as well as the duration of voicing during the final consonant closure, are reliable predictors of the voicing characteristic of the final consonant (3) other measures, such as syllable duration, final consonant duration, and vowel duration plus final consonant duration, are not as reliable as vowel duration and closure voicing duration as predictors of final consonant voicing; (4) the three- and four-year olds did not produce significantly different vowel or closure voicing duration.

Introduction

Studies of vowel production by adults have shown that vowels preceding voiced consonants in English are of greater duration than those preceding voiceless consonants (Rositzke, 1939; Heffner, 1941; Belasco, 1953; Peterson & Lehiste, 1960). That is, the vowel of /bid/ is of greater duration than that of /bit/, and that of b/z/ greater than that of /b/s/. These durational differences have been shown to be cues to the voicing characteristic of final consonants in synthetic speech (Denes, 1955; Raphael, 1972). Although it is not yet clear whether these differences in vowel duration have a physiological basis, or are entirely learned behavior, they appear to be a robust phenomena in English. Investigators report ratios from 1·25 : 1 to 2·3 : 1 between the averaged durations of vowels in opposing voicing contexts (Raphael, 1971).

*Also at Haskins Laboratories, New Haven, Connecticut 06510, U.S.A.

0095–4470/80/030335+07 $02.00/0

The purpose of the present study was to investigate the differences in duration between vowels preceding voiced and voiceless stops and fricatives in the speech of three- and four-year olds. Specifically, answers were sought for the following questions:

(1) Do young children systematically produce vowels of different durations before word-final voicing oppositions?

(2) If such systematic differences occur, are they similar in magnitude to those produced by adult speakers?

(3) Is there systematic variation in the duration of voicing during final consonant closure?

(4) How well do vowel duration; closure voicing and other durational factors (e.g. closure duration) predict, individually and in combination, the intended voicing characteristics of a word-final consonant?

(5) Are there age-related differences between three- and four-year olds in the above measures?

A number of investigators (Naeser, 1970; DiSimoni, 1972; Smith, 1978) have reported that children of the same ages as those studied here produce vowels of greater duration before voiced than before voiceless consonants. These differences are often reported to be of the same magnitude as those produced by adult speakers. Observations of final consonants (Velten, 1943; Naeser, 1970; Hodson, 1975) suggest that younger children do not distinguish cognates by the differential use of vocal pulsing. In fact, the appearance of vocal pulsing in word-final position has been said to follow the acquisition of vowel duration differences in a developmental sequence. It should be noted, however, that closure voicing during word-final consonant articulation is not consistently found in adult speech (Jones, 1950; Gimson, 1962; Lisker & Abramson, 1964). Thus it is possible that children acquire the ability to differentiate word-final cognate consonants on the basis of closure voicing only to exercise the ability inconsistently, or according to some as yet unformulated contextual rule(s) as they mature.

Method

The 20 informants were 9 three-year old and 11 four-year old children. Seventeen of the children were enrolled in a nursery school in the Bronx, New York City. The remaining three subjects were children of students and of faculty members of Herbert H. Lehman College of the City University of New York. The age range was from three years and three months to four years and eight months.

Seven minimal pairs were elicited from the children and recorded on audio tape. The seven minimal pairs were: (1) *rope–robe,* (2) *Bert–bird,* (3) *tight–tied,* (4) *pick–pig,* (5) *peck–peg,* (6) *leaf–leave,* and (7) *loose–lose.* All of the English stop contrast and the two most common fricative contrasts are contained in the list of utterances. The vowel inventory was representative of all categories of tongue height, tongue advancement and tongue "tension".

The experimenters attempted to elicit each word without prompting – that is, without saying the words themselves – as the last element in a syntagmatic response frame or as the response to a picture or object shown, or action performed. Each child was told that he would be playing a word-guessing game and was given six practice items which have not been incorporated into the data. Once the subject seemed to understand the task, the elicitation of the minimal pairs listed above began. With the exception of the second pair on the list (*Bert–bird*), no minimal pair items were elicited consecutively.

If a child did not respond, or if the child's response was not the target word, then the experimenter spoke the word aloud and said that he would ask again later to see if the child remembered it. Fairly often the child did, and the word was recorded. Such responses have been noted as delayed imitations, since the word was not said by the experimenter immediately before the child spoke it. If the child did not subsequently remember the word, then the investigator spoke it and asked the child to repeat it. We have noted such responses as immediate imitations. Often the word was elicited still later in the session, and thus several items were recorded first as immediate imitations and then as delayed imitations. Any "correct" response was immediately followed by a request for repetition.

Of the more than 600 tokens which were initially analyzed, 18% were immediate imitations, 21% were delayed imitations, and 61% were unprompted responses. The necessity for prompting varied greatly from word to word. For example, since all of our subjects watch Sesame Street, *Bird* (from Big Bird) and *Bert* (from Ernie and Bert) never had to be prompted. On the other hand, very few of the subjects produced the verbs *leave* and *peck* without prompting. No systematic differences have been found among the data derived from the unprompted responses, delayed imitations, or immediate imitations. Nevertheless, all of the immediate imitations have been eliminated from the data except for those which were the only instances of a response type for a given subject.

Since we asked for an immediate repetition of any desired response, most of the data is derived from two or more tokens of each response type per child. Those data derived from only one token of a response type generally resulted from the elimination of other tokens because of extraneous background noise which rendered the spectrograms "unreadable." or from various other causes such as sudden upward leaps in fundamental frequency which virtually transformed spectrograms from wide to narrow band and thus removed the formant information which is essential for segmenting and measuring.

Wideband spectrograms were made of each utterance using a Kay Sonagraph, model 6061B. The tapes were played into the Sonagraph at half-speed in order to lower the frequencies of the children's speech, thus making the presence, initiation, and termination of fundamental frequency more easily discernible. This also facilitated the durational measurements by expanding the time scale of the spectrograms.

The following measures, estimated to the nearest 5 ms, were derived from the spectrograms for each token:

(1) The duration of the vowel.
(2) The duration of the voicing during the final consonant closure.
(3) The duration of the closure for the final consonant.
(4) The total duration of the vowel plus the final consonant.
(5) The syllable duration.
(6) The total duration of voicing during the syllable.
(7) The total duration of voicing during the vowel and final consonant closure, which in the case of the stop-initiated syllables was the same as the total duration of voicing during the syllable.

We will not dwell at length on the difficulties to be encountered in segmenting and measuring durations of speech sounds in their spectrographic representations. It is clear that vowel duration differences of the magnitude which we have most often found are easily discernible on spectrograms. Specifying durations, however, is considerably more difficult than simply noting that they are there. This is because of the segmentation problem, and the degree of difficulty depends to a great extent on the utterance in question. A pair of the type *tight–tied* presents minimal difficulty: the burst-releases for the stops are clear indications of both the beginning and end of the syllable; the onset and offset of formants and of fundamental frequency are generally easily visible and adequately define the limits of the vowel and of the voicing during the consonant closure. On the other hand, a pair of the type *leaf–leave* presents some serious difficulties, among them being the segmentation of the initial consonant from the vowel, and definition of the limits of the final consonant, especially its termination in low-intensity noise, When uncertainty ran too high, utterances were simply eliminated from the corpus. Thus 25% of the originally recorded and spectrographicly displayed tokens were not measured. Otherwise, correspondences were sought between all the tokens of both members of a minimal pair, and some recurring acoustic event was used as a landmark for segmentation. For example, voicing during final consonant closure was measured up to the first break in the regular pulsing of the vocal folds as delineated by the low frequency vertical striations on the wideband spectrogram. This landmark was used even though there were many examples of sporadic voicing after the break and occasionally throughout the final closure. Since the intensity of closure voicing tends to decrease over time, it is not at all certain whether the vocal pulses are audible, even before the voicing break used as a landmark. If any of the pulsing is audible, however, it seems reasonable to assume that it will occur before the transglottal pressure differential has diminished sufficiently to introduce a hiatus in vocal fold vibration.

Results and discussion

The average vowel durations for both members of each minimal pair are shown in Table I. In 95% of the tokens the vowels are of greater duration when preceding voiced consonants than when preceding voiceless consonants. Of the 20 subjects, 14 showed no reversals of vowel duration for any of the oppositions. That is, the vowel preceding a voiceless consonant for these subjects was never as long or longer than the one preceding the voiced cognate consonant. Two of the utterance pairs (*rope–robe* and *pick–pig*) showed no reversals for any subjects and only one pair (*peck–peg*) provided as many as two reversals. Of those six subjects who produced vowels of equal or greater duration before voiceless consonants only one did so for two minimal pairs (*Bert–bird* and *loose–lose*), the others reversing in one pair only.

The durational differences found were similar in magnitude to those produced by adults. The range of adult vowel duration ratios from 1·25 : 1 to 2·3 : 1 contains fully 90% of the tokens in our data. Of those outside this range almost all were greater than 2·3 : 1. These results for vowel duration are in general agreement with those of Naeser (1970), DiSimoni (1972) and Smith (1978).

Table I Averaged durations and ratios of vowels and of voicing during consonant closure of the members of each minimal pair

Minimal pair	Average vowel duration (ms)	Ratio	Average closure voicing duration (ms)	Ratio
Rope Robe	129·7 209·4	1·6 : 1	50·0 112·5	2·3 : 1
Bert Bird	227·9 321·5	1·4 : 1	48·7 80·8	1·7 : 1
Tight Tied	221·8 324·5	1·5 : 1	30·5 75·3	2·5 : 1
Pick Pig	99·8 196·3	2·0 : 1	20·3 48·2	2·4 : 1
Peck Peg	131·6 198·0	1·5 : 1	24·5 86·3	3·5 : 1
Leaf Leave	150·8 247·9	1·6 : 1	48·3 89·7	1·9 : 1
Loose Lose	136·3 231·8	1·7 : 1	57·5 121·8	2·1 : 1

We now turn to the other durational measures to see how well they predict the voicing class of word-final consonants. Table I displays the averaged duration of closure voicing for voiced and voiceless consonants for each minimal pair. For each contrast the mean duration is greater for the voiced consonants. In 90% of the contrasting tokens the closure voicing is of greater duration for final voiced consonants than for the voiceless ones. Of 132 contrasts there are only 13 reversals in the duration of closure voicing. We observe that these data for closure voicing contradict those of Hodson (1975) which indicated that children as old as four years do not produce voiced obstruents in final position.

It is interesting to note that (in Table I) the ratios of the closure voicing durations from the voiced to the voiceless contexts are consistently greater than those of the averaged vowel durations for each minimal pair. If we assume perceptual significance for these temporal features, then the effect of the difference in ratios would be to maximize the salience of the potential perceptual cue (closure voicing) with the lesser duration.

Although neither vowel duration nor closure voicing perfectly predicted the intended voicing characteristic of the word-final consonants the combination of the two measures did. That is, the sum of the durations of vowel and closure voicing for each subject was always greater for the member of a minimal pair ending in a voiced consonant than for the member ending in a voiceless consonant.

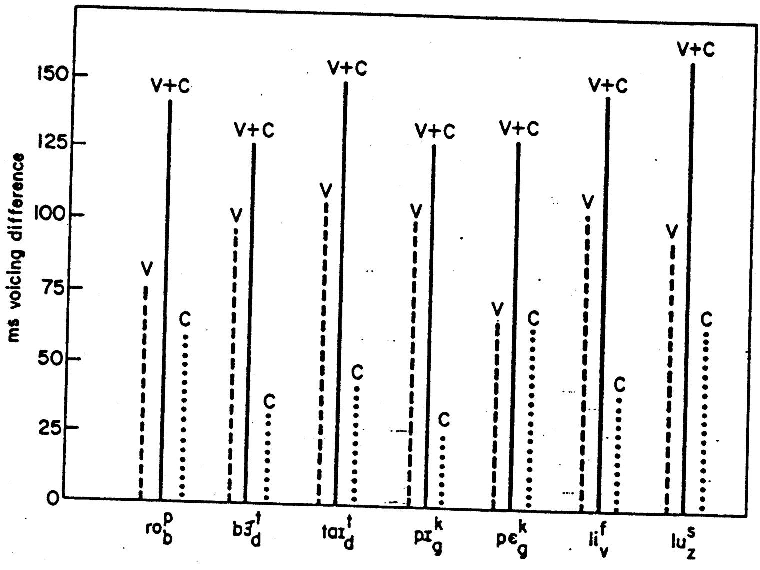

Figure 1 Averaged durational differences between vowels, closure voicing, and the sums of both measures.

Figure 1 displays the magnitude of the averaged differences between the durations of vowels, closure voicing and both measures taken together. The averaged differences for the combined measures for each of the contrasts fall between 124 and 160 ms. It is tempting on the basis of these data to speculate that some small range of overall differences is being aimed at by speakers. The range of overall difference for vowel plus closure voicing is less than 36 ms. However the way in which the total difference is divided between vowel and closure voicing varies considerably from one utterance pair to the next. In *peck–peg* for example the total difference comprises almost equal differences between vowel duration and closure voicing duration. In *pick–pig* on the other hand, more than 4/5 of the total difference is supplied by the difference between the durations of the vowels.

With one exception, to be discussed below, none of the other spectrographic measurements predicted the voicing characteristic of the final consonant as well as either vowel duration or closure voicing duration (or, of course, as well as the summed measures). The closure duration of the final consonant was greater for the voiceless cases in 71% of the tokens. This finding is in general accord with studies of final-stop consonants for adult speakers of Breton (Falc'hun, 1951), Swedish (Löfqvist, 1975) and with Smith's (1978) findings for adult and two- and four-year old English speakers. We find, however, a smaller ratio of averaged voiced to voiceless closure durations than did Smith for his two- and four-year old subjects. Our data was further marked by inconsistency both within and across subjects with regard to this measure, both for stop consonant closure and fricative durations. The sum of vowel duration and final consonant closure duration was greater for 79% of the syllables ending in voiced consonants. This last result is to be expected, of course, since the effectiveness of vowel duration as a predictor is reduced by that of closure duration which was most often greater in the case of the voiceless consonants. The total syllable duration was greater for the voiced context in 65% of the tokens. The predictive power of this measure suffers because of the variability of the durations of the consonants, especially those in initial position which varied asystematically with regard to the voicing class of the final consonants.

The only other spectrographic measure that predicted the intended voicing class of final consonants as well as did vowel duration and closure voicing was the total duration of voicing in the syllable. This measure predicted final consonant voicing characteristics in 97% of the tokens. The strength of the measure thus exceeds all but that of the perfect predictive power of the summed vowel and closure voicing durations. It falls short of perfection because of the adverse effect of the asystematic variation in the duration of voicing in the initial resonant consonants. In the cases of stop-initiated syllables, this measure was identical with that of the summed durations of vowel and closure voicing.

Finally, we consider age-related differences in the productions of the three– and four–year olds. The average differences between durations of vowels, closure voicing, and the two measures combined are shown in Table II. Although large, the age-related differences are not statistically significant.

Table II Average differences in ms for vowel duration, final-consonant closure-voicing duration, and both measures combined

		Vowel	Closure	Vowel + closure
Three-year olds	ms	83·2	36·8	120·0
	percent	69·4	30·6	100·0
Four-year olds	ms	98·6	55·6	154·2
	percent	63·9	36·1	100·0
All subjects	ms	90·2	47·8	138·0
	percent	65·4	34·6	100·0

Differences are expressed as percentages of the combined measure

By age three, then, children produce differences in vowel duration before voiced and voiceless final consonants which are of a magnitude similar to that found in adults' productions. Further, our subjects evidenced closure voicing during final consonant articulation with greater consistency than that reported for adults, and in such a way as to differentiate cognates on the basis of acoustic measurements. Research with younger children than those studied here will be necessary in order to reveal when vowel duration and closure voicing differences are first manifest and how they develop during the early years of language use.

This work was supported in part by NIH Grant NS13167, and in part by NICHD Grant HD-01994 to Haskins Laboratories.

References

Belasco, S. (1953). Influence of force of articulation of consonants on vowel duration. *Journal of the Acoustical Society of America* 25, 1015–16.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27, 761–764.

DiSimoni, F.G. (1972). An exploratory study of articulatory behavior in three, six and nine-year-old speakers, with particular reference to the effects of various phonetic environments upon phoneme duration. Doctoral Dissertation, Temple University. (Unpublished).

Falc'hun, F. (1951). *Le Système Consonantique du Breton avec une Comparative de Phonétique Expérimentale.* Rennes: Impremeries. Reunies.

Gimson, A. C. (1962). *An Introduction to the Pronunciation of English.* London: Edward Arnold Ltd.

Heffner, R-M. S. (1941) Notes on the length of vowels (IV). *American Speech* 16, 204–207.

Hodson, B. W. (1975). Aspects of phonological performance of 4-year-olds. Doctoral Dissertation, University of Illinois. (Unpublished).

Jones, D. (1950). *The Phoneme: Its Nature and Use.* Cambridge: Cambridge University Press.

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

Löfqvist, A. (1975). A study of subglottal pressure during the production of Swedish stops. *Journal of Phonetics* 3, 175–189.

Naeser, M. (1970). The American child's acquisition of differential vowel duration. The *Wisconsin Research and Development Center for Cognitive Learning, Technical Report* no. 144.

Peterson, G. & Lehiste, I. (1960) Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America* 51, 1296–1303.

Raphael, L. J. (1971). Vowel duration as a cue to the perceptual separation of cognate sounds in American English. Full volume supplement to *Haskins Laboratories Status Report on Speech Research.*

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America* 51, 1296–1303.

Rositzke, H. (1939). Vowel length in general American speech. *Language* 15, 99–109.

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

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