Development of [ɹ] in young, Midwestern, American children

Richard S. McGowan
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420-2002

Susan Nittrouer
Center for Persons with Disabilities, Utah State University, Logan, Utah 84322-6885

Carol J. Manning
Boys Town National Research Hospital, Omaha, Nebraska 68131

Beginning at the age of about 14 months, eight children who lived in a rhotic dialect region of the United States were recorded approximately every 2 months interacting with their parents. All were recorded until at least the age of 26 months, and some until the age of 31 months. Acoustic analyses of speech samples indicated that these young children acquired [ɹ] production ability at different ages for [ɹ]'s in different syllable positions. The children, as a group, had started to produce postvocalic and syllabic [ɹ] in an adult-like manner by the end of the recording sessions, but were not yet showing evidence of having acquired prevocalic [ɹ]. Articulatory limitations of young children are posited as a cause for the difference in development of [ɹ] according to syllable position. Specifically, it is speculated that adult-like prevocalic [ɹ] production requires two lingual constrictions: one in the mouth, and the other in the pharynx, while postvocalic and syllabic [ɹ] requires only one oral constriction. Two lingual constrictions could be difficult for young children to produce. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1642624]

I. INTRODUCTION

A. Phonetics and acoustics of /r/

In rhotic dialects of American English, the /r/ phoneme is pronounced as an approximant [ɹ], and it is notoriously difficult for American children to learn to produce.[1] Sander (1972) reported that the median age for acquisition of /r/ for American children was 3 years, and it was not until age 6 years that 90% of children produced /r/ correctly. Smit *et al.* (1990), in their study of 3- to 9-year-old children from Iowa and Nebraska, reported that 90% of the children had attained correct /r/ production by 8 years of age.

In adults' productions of prevocalic [ɹ], the most clearly defining acoustic property is a very low third formant frequency, $F3$. It often dips below 2000 Hz, which is well below its value for a neutral vowel. Another acoustic correlate of [ɹ] is that $F3$ is generally close to $F2$. That is, the value of $F3 - F2$ is smaller in [ɹ] than it is for a neutral vowel (Lehiste, 1962; Delattre and Freeman, 1968). However, variation exists in the precise acoustic properties of [ɹ] depending on whether it is prevocalic, postvocalic, or syllabic (Lehiste, 1962; Delattre and Freeman, 1968; Olive, Greenwood, and Coleman, 1993). The word "right" contains a prevocalic [ɹ]; the word "car" possesses a postvocalic [ɹ]; the name "Burt" has a medial syllabic [ɹ]; and the word "doctor" has a final syllabic [ɹ]. In general, the first three formant frequencies are not as low in postvocalic and syllabic [ɹ] relative to their values in a neutral vowel, as they are in prevocalic [ɹ] in a stressed syllable. However, $F3$ is lower and $F3 - F2$ smaller than their expected values in neutral vowels for all of these variations of [ɹ]. The properties of low $F3$ and smaller $F3 - F2$ are what we call properties of [ɹ]-*ness*. Thus, the previous literature indicates that details of the formant structure for an [ɹ] depends on its syllable position.

The literature on young American children's production of /r/ is sparse and often limited to word-initial or prevocalic /r/. Dalston (1975) studied the formant frequencies of word initial /r/, /w/, and /l/ in adults and children who were 3 to 4 years of age. He confirmed that a relatively low $F3$ occurred when the children produced word-initial /r/ correctly as [ɹ], and that this $F3$, with a mean of approximately 2500 Hz, helped to distinguish /r/ from /w/ and /l/, which both had mean $F3$'s of approximately 3500 Hz. Also, scatter plots of $F3/F1$ versus $F2/F1$ for the adults and children revealed that the children produced /r/ and /w/ with more overlap in the $F2/F1$ parameter than did the adults. This overlap could be a contributing factor to adults perceiving /w/ when word-initial, prevocalic /r/ is produced incorrectly by children (e.g., Smit *et al.*, 1990; Smit, 1993; Shriberg and Kent, 1995). In a phonetic study of segment acquisition in children 15 to 24 months of age, Stoel-Gammon (1985) found that word-final /r/ was acquired before word-initial /r/. This corroborated other, earlier studies of American English speech sound acquisition. The phonetic studies of Smit *et al.* (1990) and Smit (1993) indicated that 2- to 5-year-old children from Iowa and Nebraska could produce intersyllabic /r/ and syllabic /r/ with substantially fewer errors than they could produce prevocalic /r/.

The protracted period of development of [ɹ] allows us to examine the development of an English speech segment. Further, the acoustic data on adult [ɹ] indicate that the way it is produced depends on syllable position. We were interested in whether young children also show differences in production of this segment depending on syllable position, as well as whether the speed with which [ɹ] developed depended on syllable position. While similar developmental trends may

exist for segments other than [ɹ], they might be more easily missed in longitudinal studies if children progress through the stages of development rapidly.

B. Articulatory considerations and articulatory–acoustic relations for /r/

The acoustic correlates of the [ɹ] segment can be described, and it is known that several different articulatory gestures can give rise to these acoustic characteristics. Yet, an aspect that provides an understanding of the development of [ɹ] is an understanding of the articulatory gestures that children use to produce /r/ and how these gestures differ across syllable position. Along with a review of the literature of adult's [ɹ] production, an example of simultaneously recorded acoustic and articulatory data from an adult talker will be examined below to develop a hypothesis about the possible articulatory gestures that could underlie the acoustic data that were collected from children. The procedures required to obtain data on the articulatory gestures involved in /r/ production are more invasive than generally considered acceptable for work with young children. While adults cannot provide much insight into children's articulatory behavior, they can provide insight into the physical relations between articulation and acoustics.

Delattre and Freeman (1968) performed an extensive study of adults' American English [ɹ] using x-ray cineradiography with simultaneous acoustic recording. Their results showed that each of their American English speakers produced stressed, prevocalic [ɹ] with two vocal-tract constrictions: one palato-velar and the other pharyngeal. There was a variety of tongue shapes, with either the tongue dorsum, tongue blade, or tongue tip providing the palato-velar constriction. Also, a variety of constriction degrees was used, and narrower constrictions were correlated with lower $F3$ frequencies (and so with smaller $F3-F2$ differences, as well). Furthermore, lip rounding was used consistently by their speakers for prevocalic /r/ in stressed syllables. This had the effect of lowering $F1$, $F2$, and $F3$ from their values in postvocalic /r/, which was not produced with lip rounding. Otherwise, Delattre and Freeman (1968) observed that the postvocalic [ɹ] had the same tongue shapes as prevocalic [ɹ], but with $F3$ not as low and $F3-F2$ not as small as for stressed, prevocalic [ɹ].

The relation between the acoustics and articulation of [ɹ] has received attention more recently. Stevens (1998) proposed a model for retroflexed [ɹ] production in which the volume under the tongue creates an acoustic side branch that gives rise to a pole–zero pair. The pole would constitute a formant that appears between the formants that are continuous with the $F2$ and $F3$ of the surrounding vowels. However, a different model of [ɹ] production, based on the three-dimensional MRI data of Alwan, Narayanan, and Haker (1997), was proposed by Espy-Wilson et al. (2000) and by Jackson, Espy-Wilson, and Boyce (2001). These investigators judged the dimensions of the sublingual cavities for the two subjects producing [ɹ] in the MRI study to be too small to account for an extra formant just above $F2$ and below $F3$, as proposed by Stevens (1998). Rather, they proposed that the cavity in front of the palatal constriction was of sufficient

size to produce a low $F3$. Whether this front cavity should be modeled as a Helmholtz resonator or as a quarter-wavelength resonator would depend on the degree of lip rounding. They also showed that the region of the oral cavity behind the tongue tip could be modeled either as a double-Helmholtz resonator or as a single-Helmholtz resonator in series with a half-wavelength resonator accounting for $F1$ and $F2$, depending on the degree of the pharyngeal constriction. Thus, in the case where both the palatal and pharyngeal constrictions are tight and lip rounding is present, the entire system behaves more or less as three coupled Helmholtz resonators, from which three low-frequency formants result. However, if the pharyngeal constriction is only moderate, it is better to model the situation as two coupled Helmholtz resonators, with the coupling at the palatal constriction. The palatal constriction produces a low $F1$ and $F3$, while the moderate pharyngeal constriction lowers $F1$ and $F2$. Espy-Wilson and Boyce (1999) reported that $F4$ is relatively low for retroflexed articulations compared to its value for nonretroflexed articulations. If $F4$ is a resonance of the cavity behind the palatal constriction, then its low value could be the result of the palatal constriction of the retroflexed [ɹ] being farther forward or shorter than for the nonretroflexed [ɹ].

Others have studied American English [ɹ] articulation from different perspectives. For example, using MRI technology Alwan et al. (1997) provided much-needed three-dimensional data on sustained [ɹ], from which acoustic tube models could be constructed. One subject, PK, produced both sustained retroflexed and nonretroflexed [ɹ] with both tight palatal and tight pharyngeal constriction. Another subject, MI, produced sustained word-initial and syllabic [ɹ] with a tight palatal constriction, but only a moderate constriction in the pharynx. Lip rounding was involved in all these productions. One of their observations was that American English retroflexed [ɹ] was actually produced by a raised, laminal tongue blade, and not a curled tongue blade. Westbury, Hashi, and Lindstrom (1998) used x-ray microbeam technology with the data from the X-Ray MicroBeam Speech Production Database, XRMB-SPD (Westbury, 1994) to describe a continuum of articulatory shapes from bunched to retroflexed articulation for prevocalic [ɹ] in a large cohort of adult subjects. (A bunched articulation is one in which the tongue body is used to make the palatal constriction.) Guenther et al. (1999) used electromagnetic articularometry to show that there was an articulatory trade in seven talkers for stressed, prevocalic [ɹ] production: As the cavity in front of the constriction became shorter, due to articulatory constraints of the phonetic context, the palatal constriction widened and/or constriction length increased, thus enabling $F3$ to remain low in differing phonetic environments. There was also a study of [ɹ] production using the very minimal technology of cotton swabs to find the position of the tongue blade during prevocalic, postvocalic, and syllabic [ɹ] production (Hagiwara, 1994). In this study [ɹ] articulations were classified as tip-up, tip-down, or blade-up, depending on whether a cotton swab through the incisors touched the underside of the tongue blade, the tongue tip, or the upper surface of the blade, respectively. All three articulation types could occur for all syllable positions. The combined result of

these experiments is that speakers use a wide variety of articulatory maneuvers to produce [ɹ], including nonretroflexed tongue bunching near the palate, as well as retroflexed configurations.

The articulatory and acoustic correlates of [ɹ] according to syllable position were of great interest to this work because we wanted to understand the articulatory–acoustic relations children use in producing /r/ in different syllable positions. To further understand the physics of the articulatory–acoustic relations in [ɹ] beyond that provided by the work already cited, we examined a subject from the XRMB-SPD, JW11, who exhibited both a retroflex and bunched prevocalic [ɹ]. In the word "right" JW11 produced [ɹ] using a retroflexed tongue blade, while he produced [ɹ] with a nonretroflexed articulation in "rag." Three tokens of the prevocalic [ɹ] in "right" and two tokens of prevocalic [ɹ] in "rag" were examined. Along with these words containing prevocalic [ɹ], we examined examples of postvocalic [ɹ] in one token of "there," three tokens of "large," and two tokens of "dormer" (first syllable), as well as final syllabic [ɹ] in two tokens of "dormer" (second syllable). These observations provided the following insights for JW11, the details to which are contained in the Appendix. Given equivalent syllable stress and phrase positions, $F1$, $F2$, and $F3$ were generally higher for postvocalic and syllabic [ɹ] than for prevocalic [ɹ] because: (1) compared to prevocalic [ɹ], there was reduced or no lip rounding for postvocalic and syllabic [ɹ], and (2) the palatal constriction for postvocalic and syllabic [ɹ] was not as tight as for prevocalic [ɹ]. It cannot be determined whether the subject produced varying degrees of pharyngeal constriction across word positions because there is no indication of tongue root or larynx position in the XRMB-SPD. However, we speculate that there was reduced or no pharyngeal constriction for postvocalic and syllabic [ɹ], while there was substantial pharyngeal constriction for prevocalic [ɹ]. In support of this speculation, it was noted that the tokens with the lowest $F2$'s for postvocalic [ɹ] were those with neighboring back vowels, so that any pharyngeal constriction for these [ɹ]s may be the result of carryover coarticulation.

Based on the observations of JW11 and of the previously published work reviewed above, the articulation of [ɹ] according to syllable position can be summarized as follows: Prevocalic [ɹ] is a consonant articulated with at least one close approximation in the palato-velar region, along with a secondary constriction in the pharyngeal region, and some degree of lip rounding. Postvocalic [ɹ] is more of an off-glide to the preceding vowel with one primary constriction target in the palatal region with little or no lip rounding. Syllabic [ɹ] is a monophthong vowel with a steady constriction that is similar to that of the constriction target of the postvocalic [ɹ].

C. The present study

The present study quantifies observations of children's /r/ production in terms of $F2$, $F3$, and their separation, $F3-F2$. Also, the formant frequency trajectories for /r/ in certain vowel contexts will be compared. Statistical analyses will include linear regression so that changes of the formant frequencies and their separation for /r/'s in different syllable position can be quantified as a function of age. For tokens in

similar age groups, rank-order statistics will help quantify the differences between each $F2$, $F3$, and $F3-F2$ as functions of /r/ syllable position. While there are some reasons to believe that there could be differences in children's perception of [ɹ] depending on syllable position and status, we focus on articulatory causes for the differences in the development of [ɹ] production. As reviewed above, the prevocalic [ɹ] is articulatorily more complex than the postvocalic [ɹ] and syllabic [ɹ]. Thus, as has been observed phonetically in previous literature (e.g., Smits, 1993), we expect prevocalic [ɹ] not to be as well developed as the [ɹ]'s in other syllable positions for young children.

II. METHOD

A. Subjects

Nine children from eastern Nebraska and western Iowa were recruited for a longitudinal study of speech production, but one child left the area before completing the study. The children recruited were typically developing children. All had normal prenatal histories, normal deliveries, and no special medical conditions. None of the children had a family member with speech, language, or hearing disorders. None of the children had any reported history of otitis media with effusion at the start of the study, and no child was treated for more than one episode while the study was being conducted.

B. Procedures

Recording sessions were started as soon as possible after each child began producing consistent phonetic forms. At the start of the recording sessions all children were about 1 year of age and had vocabularies of fewer than 10 words. Children were recorded approximately every 2 months. However, parents were asked to contact the laboratory staff if they noticed what appeared to be a particularly rapid proliferation of new words, or when they noticed their child was starting to combine words. Recording sessions were discontinued when a child was consistently using sentences of three or more words, with some function words.

Children were recorded in the same sound-treated chamber at each session. Sessions were 20 min long. The child sat in a highchair at a table, with one parent across the table. The same set of toys was available for play at each session, and consisted of such things as small stuffed dolls, foam puzzles, and cloth books. All toys used in these sessions were soft to minimize extraneous noises that might interfere with speech recording. The toys were not chosen to elicit any particular response from the children. Parents were instructed to play with their children, trying to elicit as much language as possible. Also, parents kept a diary of new vocabulary items (at the younger ages) and new sentence structures (at the older ages) that they heard at home.

Recordings were obtained using an AKG C-535EB microphone, a Shure model M268 mixer, and a Nakamichi MR-2 cassette deck. This system provided a flat-frequency response out to 20 kHz. The microphone was suspended roughly 9 in. above the child's mouth. It was suspended rather than table mounted because pilot work showed that children habituated to its presence more rapidly that way.
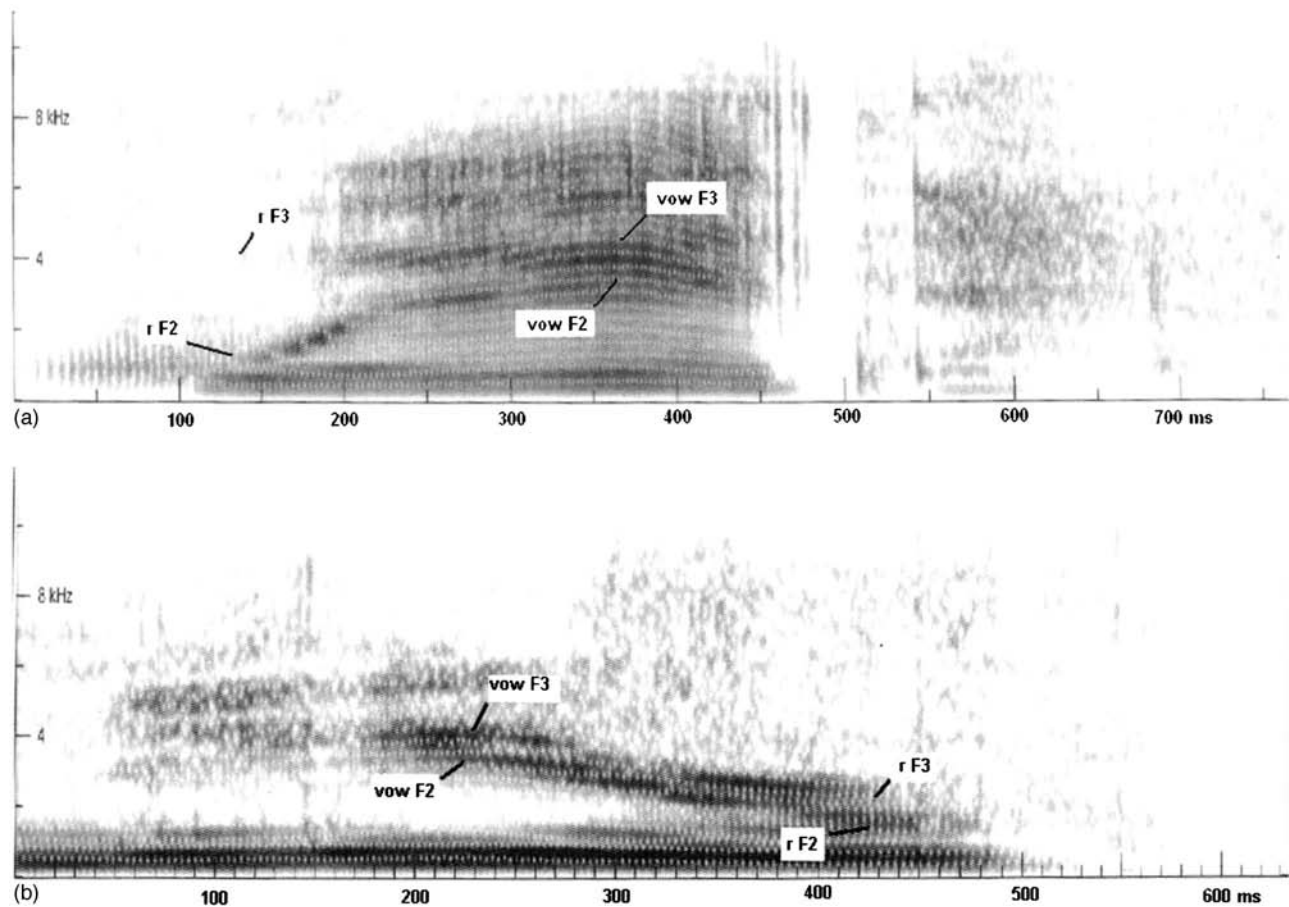
FIG. 1. (a) A spectrogram of a subject's production of the word "red." The $F2$ and $F3$ in the /r/ and in the /ɛ/ are indicated, as measured using spectral cross sections and the spectrogram. (b) A spectrogram of a subject's production of the word "here." The $F2$ and $F3$ in the /r/ and in the /i/ are indicated, as measured using spectral cross sections and the spectrogram.

These recordings were subsequently digitized with a Soundblaster A/D card using SPEECH STATION II software at a 22.05-kHz sampling rate.

The recordings of each child were examined and analyzed. For each child, the final recording was examined first, the penultimate recording second, and so on until the first recording of that child was examined. This allowed the examiner to acclimate to each child's speech starting with what should have been the most intelligible sample. Utterances with words that would have contained an [ɹ] or syllabic [ɹ] if spoken by an adult with a rhotic dialect were extracted. Conversational context helped in the identification of these utterances. Also, the children often repeated a single lexical item while engaged in play. The one exception to the way utterances were selected for analysis was from a child who produced the word "bakery" alternately as "brakery" or "bwakery," and that inserted /r/ was analyzed. There were no other instances of such /r/ insertions.

Using a spectral analysis program, SPEECH STATION II, the $F2$ and $F3$ of each word were measured in the temporal region most clearly affiliated with the /r/ phoneme, as well as in the middle of the neighboring vowel, using DFT spectral cross sections and spectrograms. Figures 1(a) and (b) show examples of spectrograms of words with prevocalic /r/ ("red") and postvocalic /r/ ("here"), respectively, as spoken by two of the children. Hamming windows of 256 samples

(11.6 ms) were used. The windows overlapped by 5.8 ms, or 128 samples. In the case of prevocalic /r/, the neighboring vowel followed the /r/ (or attempted /r/). For postvocalic /r/, the neighboring vowel preceded the /r/. Intersyllabic /r/ was counted as prevocalic, although there were not many of these. Formant measurements in the diphthongs /aɪ/ and /eɪ/, which only appeared with prevocalic /r/, were taken in the off-glide at the most steady portions and before transition to closing consonants. For prevocalic /r/, $F2$ and $F3$ were measured at the $F2$ minimum. In the case of postvocalic /r/, either the local minimum or local maximum value of $F2$ was chosen as the measurement time, depending on whether the $F2$ trajectory from the preceding vowel was falling or rising. Measurement of medial syllabic /r/, as in "Burt," was done the same way as for postvocalic /r/, except that formants were measured toward the beginning and toward the end of syllabic /r/, while avoiding the surrounding consonantal transitions. This was also the measurement method applied to word-final, unstressed, syllabic /r/, as in "keeper." The two types of syllabic /r/'s, medial and final, were analyzed separately because it was not clear at the outset whether word-final syllabic /r/ would behave more like a vowel-and-postvocalic-/r/ or more like a medial syllabic /r/. Only those words for which $F2$ or $F3$ data could be measured were included in the analysis, so that words which were too faint or produced in a scream were excluded from the analysis.

Further, $F2$ was required to be between 1000 and 3500 Hz and $F3$ between 2000 and 5100 Hz. The values close to these lower bounds were expected only during the supposed /r/ segment. We were careful to view the spectrogram and spectral cross section simultaneously while measuring formant frequencies. We were aware of the possibility of subglottal formants if the voice was breathy and the presence of nasal poles in nasal consonant context. None of the children exhibited hypernasality. The measured formants needed to be continuous between the /r/ and the neighboring vowel. This minimized the possibility of misidentifying a nasal or subglottal formant as a resonance of the supraglottal, non-nasal portion of the vocal tract. The results will be presented in terms of formant frequencies, $F2$ and $F3$, in /r/ and formant separation, $F3 - F2$ in /r/. Also, when specific vowel contexts are considered, the differences of $F2$ in /r/ and in the neighboring vowel, and of $F3 - F2$ in /r/ and in the neighboring vowel will be discussed.

The first and fourth formant frequencies were not recorded for the children because they simply could not be measured reliably in a sufficient number of tokens. $F1$ was frequently without acoustic energy due to children's high fundamental frequencies: often the fundamental frequency was higher than $F1$; $F4$ was often too faint due to the steep spectral tilt for many of the children's productions. The more reliably measured $F2$ and $F3$ turned out to be sufficiently indicative of differences in /r/ production according to syllable position.

Statistical analyses consisted of linear regression and rank-order analysis. Comparisons of linear regression slopes for these measures across age, grouped according to syllable position, allowed comparisons of the rates of change with age in the formants and formant separations. Because vocal-tract size changes, we can expect that the formant frequencies $F2$ and $F3$ will decrease with age, so that it was best to compare slopes of the regression lines for /r/ in different syllable position. We treated formant separation in a similar way. Because of the variability in the data and the limited number of data points, we accepted an 80% confidence in slope differences as significant. In another kind of analysis an acoustic quantity, such as $F3$, was compared between syllable positions within an age group. In this kind of comparison, rank-order statistics was used because formant frequencies were often distributed non-normally and because they allowed two groups of tokens with relatively low sample sizes to be compared. The probability that two random variables (e.g., $F3$ for /r/ in prevocalic position and $F3$ for /r/ in postvocalic position) possess different probability distributions, even when those distributions are not known, can be tested using rank-order statistics. The particular statistic employed here was the *Wilcoxon* statistic (Bickel and Docksum, 1977, pp. 344–355). In fact, if the hypothesis that the two random variables have the same probability distribution is rejected then one of them is *stochastically greater* than the other. Symbolically, distribution F is stochastically greater than distribution G, if the probability that random X with distribution F is greater than any given t, is greater than the probability of a random variable Y with distribution G being greater than t (Bickel and Docksum, 1977, pp. 344–

355; Fisz, 1963, pp. 451–456). In all comparisons using rank-order statistics, the sample sizes were at least ten.

III. RESULTS

The data are presented in categories of age, in months, at recording. Months are grouped by two, except for month 19, when no recording of the eight subjects was made. Using age categories of 2-month duration ensured that enough subjects and recording sessions would be included in each age category. Table I shows the number of samples of prevocalic, postvocalic, medial syllabic, and final syllabic /r/ extracted for analysis for each subject in each age category. Except for one case, this meant that a subject could be included only once in each age category, although not all subjects were included in all age categories. The exception was BT, who was recorded at both 26 and 27 months, and, hence, the number of recording sessions is one more than the number of speakers for this age category.

A. Numbers of words with /r/ or syllabic /r/

Figure 2 shows the mean number of words per recording session, with standard deviations, for analyzed words that had, according to adults' productions, prevocalic /r/, postvocalic /r/, medial syllabic /r/, and final syllabic /r/. These numbers do not indicate whether the children produced perceptually acceptable versions of /r/. The mean number of prevocalic and postvocalic /r/'s per session generally increased across sessions. For postvocalic /r/, the mean incidence increased rapidly over the first three recording sessions, and then remained stable until the last recording session, when there is a dramatic increase again. For prevocalic /r/, the mean incidence remained low until the fourth recording session, when it increased sharply.

There are some important qualifications that should be made about the kinds of tokens that were elicited in the recording sessions. There was no effort made in recording sessions to elicit the same utterances at each session, so there might simply have been variability in the numbers of words with /r/ due to what the parent–child dyad was discussing. For instance, while the rate of final syllabic /r/ word productions increased with age, this was not true for the medial syllabic /r/ production. There was a peak in the latter in the 24–25-month category, with a steady decrease after this age (Fig. 2). A closer look at the data revealed that the frequent producers of medial syllabic /r/'s made many references to the ''Burt'' and ''Ernie'' dolls in the recording room. As a result, it might be the case that rate of medial syllabic /r/ production was actually stable across this period of early language acquisition. Finally, children were dismissed from further recording sessions when they began to routinely produce short (i.e., three-word) sentences. Thus, only the children developing speech at the slowest (normal) rate were included in the last recording session.

B. Children's formant frequencies according to syllable position

Initially, averages of $F2$, $F3$, and $F3 - F2$ in /r/ as a function of syllable position and age category were exam-

TABLE I. Table of age categories for which /r/ and syllabic /r/ data were collected for each subject. The numbers in each cell (left to right, top to bottom) represent the number of (1) prevocalic /r/ tokens; (2) postvocalic /r/ tokens; (3) medial syllabic /r/ tokens; and (4) final syllabic /r/ tokens. The bracketed pairs for BT at 26–27 months denote each of these quantities recorded in two separate sessions. "X" is used to indicate that the subject was not recorded for a particular age category.

	15–16	17–18	20–21	22–23	24–25	26–27	28–29	30–31
AN	0, 0, 0, 0	0, 0, 0, 0	3, 6, 1, 0	11, 11, 4, 0	11, 22, 19, 3	X	X	X
BT	0, 1, 0, 0	0, 1, 0, 0	1, 0, 0, 1	X	X	(9,22), (7,7), (8,2), (0,4)	X	X
CK	0, 0, 0, 0	0, 0, 0, 0	0, 0, 0, 0	2, 12, 2, 0	1, 4, 4, 0	X	2, 24, 9, 1	17, 30, 3, 3
DF	0, 0, 0, 0	0, 0, 0, 0	4, 3, 0, 0	9, 7, 3, 1	8, 20, 8, 3	5, 7, 3, 6	21, 1, 3, 6	14, 14, 0, 4
LG	0, 0, 0, 0	0, 0, 0, 0	0, 0, 0, 0	0, 10, 2, 3	9, 16, 4, 7	X	10, 11, 1, 3	21, 20, 7, 19
MS	0, 0, 0, 0	0, 0, 0, 0	2, 3, 1, 0	X	0, 2, 2, 1	6, 8, 2, 0	X	X
MST	0, 0, 0, 0	3, 0, 0, 0	9, 4, 16, 3	X	15, 14, 20, 9	11, 16, 17, 5	13, 11, 7, 10	X
RF	0, 0, 0, 0	0, 0, 0, 0	2, 3, 0, 0	2, 3, 6, 2	0, 0, 3, 0	25, 15, 1, 5	22, 8, 0, 12	8, 9, 0, 2
Total number of subjects	8	8	8	5	7	6	5	4

ined. Thus, the data from different speakers and different vowel environments were pooled for the analyses in this section. For both prevocalic and postvocalic /r/ [Figs. 3(a) and (b)] there was a slight tendency for $F2$ and $F3$ to decrease with age. For $F2$ the trend appeared to be stronger for prevocalic /r/ than for postvocalic /r/: Between 26 and 31 months the mean $F2$'s for prevocalic /r/ was between 1600 and 1900 Hz, while for postvocalic /r/ the mean $F2$ was between 1900 and 2400 Hz. For $F3$ the trend appeared stronger for postvocalic /r/ than it was for prevocalic /r/: Between 26 and 31 months the mean $F3$ for postvocalic /r/ was about 3200 Hz, while it was about 3700 Hz for prevocalic /r/. The

result of these trends was a faster decrease in mean $F3-F2$ for postvocalic /r/ than for prevocalic /r/ with age category. There did not appear to be a consistent change with age category in mean $F3-F2$ for prevocalic /r/: This quantity stayed close to 2000 Hz throughout the age categories except for 14–15 months, where it was even higher. On the other hand, there was a general downward trend in this quantity with age for postvocalic /r/, so that by the 28–29-month and 30–31-month categories mean $F3-F2$ was closer to 1000 Hz. A 1000-Hz difference in frequencies is less than would be expected for a neutral vowel for ages 1 to 2 1/2 years.[2]

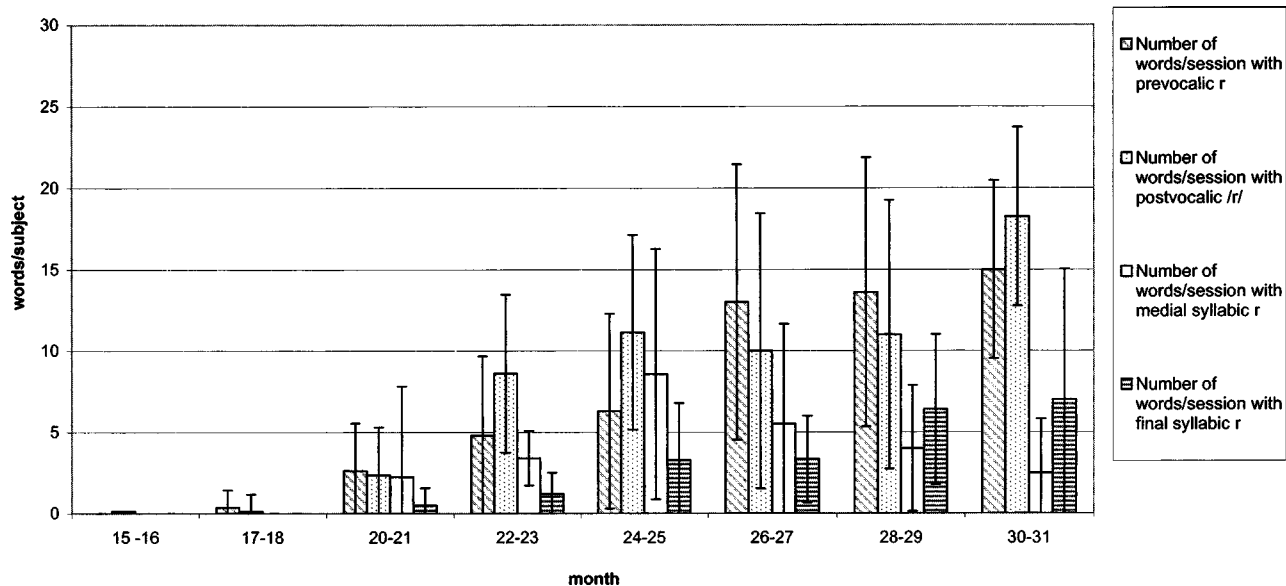Linear regression was performed on $F2$, $F3$, and

FIG. 2. Mean numbers and standard deviations of analyzed words per recording session as a function of age category. Words either possess prevocalic /r/, postvocalic /r/, medial syllabic /r/, or final syllabic /r/.
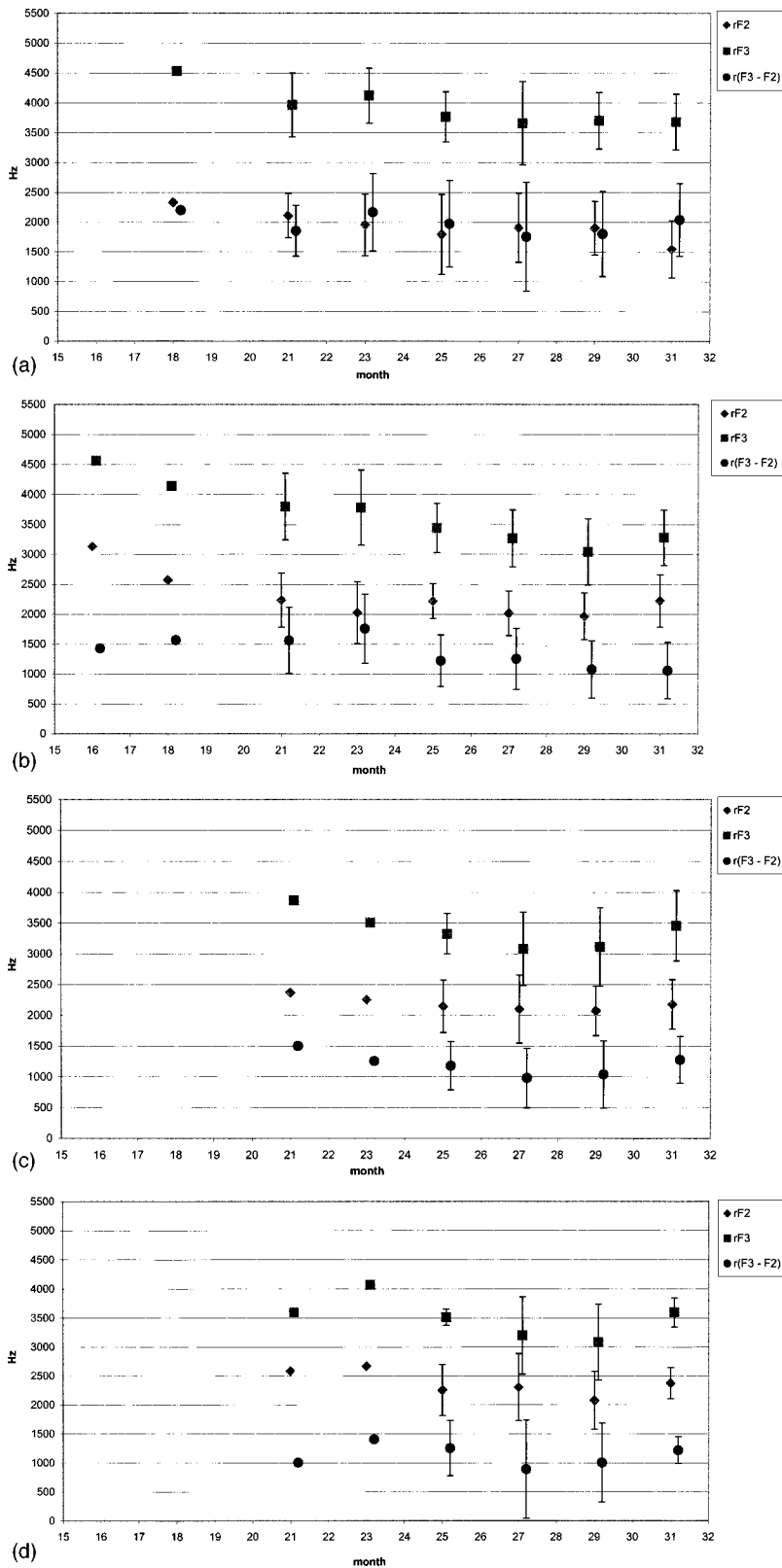
FIG. 3. (a) Means and standard deviations of $F2$, $F3$, and $F3-F2$ in prevocalic /r/ for the children subjects as a function of age category. (b) Means and standard deviations of $F2$, $F3$, and $F3-F2$ in postvocalic /r/ for the children subjects as a function of age category. (c) Means and standard deviations of $F2$, $F3$, and $F3-F2$ in medial syllabic /r/ for the children subjects as a function of age category. (d) Means and standard deviations of $F2$, $F3$, and $F3-F2$ in final syllabic /r/ for the children subjects as a function of age category.

$F3-F2$ for prevocalic, postvocalic, medial syllabic, and final syllabic /r/ versus age, in months. The slopes of the regression lines and their 80% confidence intervals are tallied in Table II. This analysis revealed that the 80% confidence ranges for the slope of the postvocalic $F3-F2$ versus age line and the prevocalic $F3-F2$ versus age line did not overlap (Table II, rows 1 and 2, column 3): The rate of decrease in $F3-F2$ was greater for postvocalic /r/ than for prevocalic /r/. Most of the difference between syllable position for the rate of change of $F3-F2$ appears to be the more rapid decline of $F2$ with age in prevocalic position than in postvocalic position, and the more rapid decline of $F3$ with age in postvocalic position than in prevocalic position (Table II, rows 1 and 2, columns 1 and 2). There was, however, overlap

TABLE II. Slopes and their 80% confidence intervals for linear regression of formant frequencies (Hz) in /r/ versus age, in months, of recording for /r/ in various syllable positions.

	$F2$	$F3$	$(F3-F2)$
Prevocalic	−35.0	−36.1	−1.1
	[−47.3,−22.6]	[−48.6,−23.6]	[−18.2,15.9]
Postvocalic	−14.7	−48.0	−33.3
	[−23.6,−5.8]	[−59.0,−37.0]	[−44.2,−22.3]
Medial syllabic	−34.2	−53.2	−19.0
	[−47.6,−17.0]	[−78.1,−28.4]	[−41.0,3.0]
Final syllabic	28.7	−15.1	−43.8
	[11.1,46.1]	[−42.8,12.6]	[−72.4,−15.2]

in the 80% confidence intervals for the prevocalic and postvocalic slopes for both $F2$ and $F3$. None of the slope comparisons showed differences at the 90% confidence level.

Figure 3(c) shows $F2$, $F3$, and $F3-F2$ as a function of age for medial syllabic /r/. Except for the 30–31-month category, there was an overall downward trend from 20–21 months in $F2$, $F3$, and $F3-F2$ for medial syllabic /r/ as a function of age category. Medial syllabic /r/ behaved similarly to postvocalic /r/ with mean $F3-F2$ at about 1500 Hz in the 20–21-month category and near 1000 Hz in the 26–27- and 28–29-month categories. Similarly, the mean $F3$ fell from about 3800 Hz at 20–21 months to 3200 Hz at 26–27 months. Table II shows that the 80% confidence intervals for the rate of decline of the formant frequencies with age for medial syllabic /r/ overlapped the 80% confidence intervals for these rates for both the prevocalic and postvocalic /r/'s (Table II, rows 1, 2, and 3).

For final syllabic /r/ [Fig. 3(d)], the mean $F3$ was between 3000 and 3600 Hz from 24–25 months through 30–31 months, without an apparent trend with age. For the same span of time, $F3-F2$ was between 800 and 1300 Hz. The regression analysis showed that $F2$, in fact, increased with age, while $F3-F2$ decreased with age (Table II, row 4). The 80% confidence intervals for the positive slope of $F2$ versus age for final syllabic /r/ did not overlap with any of the 80% confidence intervals for the other syllable positions. While the negative slope for $F3-F2$ in final syllabic /r/ possessed the largest absolute value of the four types of /r/, the 80% confidence interval was quite broad.

Wilcoxon statistics were employed to characterize the differences in the distributions of the formant frequencies between types of /r/ in the 30–31-month age category. Table III shows the probability: (1) that the $F2$'s for postvocalic, medial vocalic, and final vocalic /r/'s possessed distributions

TABLE III. The probability, to the nearest 0.001: (1) that the $F2$'s for postvocalic, medial vocalic, and final vocalic /r/'s possess distributions stochastically greater than that for prevocalic /r/'s, and (2) that the $F3$'s and $(F3-F2)$'s for postvocalic, medial vocalic, and final vocalic /r/'s possess distributions stochastically less than those for prevocalic /r/'s in the 30–31-month age category.

Prevocalic versus	$F2$	$F3$	$(F3-F2)$
Postvocalic	>0.999	>0.999	>0.999
Medial syllabic	0.999	0.599	0.985
Final syllabic	>0.999	>0.999	>0.999

stochastically greater than that for $F2$'s in prevocalic /r/'s, and (2) that the $F3$'s and $(F3-F2)$'s for postvocalic, medial vocalic, and final vocalic /r/'s possessed distributions stochastically less than those for prevocalic /r/'s in the 30–31-month age category. The $F2$'s for postvocalic, medial syllabic, and final syllabic /r/ all possessed distributions stochastically greater than that for prevocalic $F2$ with a very high probability (Table III, column 1). Further, the distributions of $F3$'s and $(F3-F2)$'s for postvocalic and final syllabic /r/'s were stochastically less than the corresponding distributions for prevocalic /r/ with a very high probability (Table III, columns 2 and 3, rows 1 and 3). On the other hand, there was almost no certainty that the distribution of $F3$ for medial syllabic /r/ was stochastically less than the distribution of $F3$ for prevocalic /r/ (Table III, column 2, row 2). This appeared to have had some effect on the probability that the distribution of $F3-F2$ for medial syllabic /r/ was stochastically less than that for prevocalic /r/ (Table III, column 3, row 2). However, it should be kept in mind that the number of tokens of medial syllabic /r/ was dwindling by the 30–31-month age category (Fig. 2).

Some of the prevocalic /r/'s appeared within consonant clusters. Hoffman, Schuckers, and Ratusnik (1977) found that certain initial stop consonants could facilitate the correct production of prevocalic and vocalic /r/ in children from about 6 years to 7 years of age. To test the effect that the alveolar and velar stop consonants had on the formant frequencies in /r/, rank-order analysis with the Wilcoxon statistic was performed comparing prevocalic /r/'s in consonant clusters with alveolar and velar stops and singleton prevocalic /r/'s in the age range from 28 to 31, months. Prevocalic /r/'s in consonant clusters with just alveolar stops and singleton prevocalic /r/'s in the age range from 28 to 31 months were also compared. In both cases, the prevocalic /r/'s in consonant clusters had distributions with $F3-F2$ stochastically less than the distribution for singleton prevocalic /r/'s, with small to moderate probabilities (0.81 for clusters with alveolar and velar stops versus singletons and 0.66 for clusters with alveolar stops versus singleton). However, the prevocalic /r/'s in consonant clusters had distributions with $F3$ stochastically greater than the distributions for singleton prevocalic /r/'s, with moderate probability (0.90 for clusters with alveolar and velar stops versus singletons and 0.91 for clusters with alveolar stops versus singleton). In these data, alveolar and velar stop consonants did not appear to promote correct prevocalic /r/ production.

C. Children's formant frequencies according to vowel context and syllable position

A more detailed analysis of the nature of prevocalic and postvocalic /r/ production can be attained when the word tokens are segregated according to both syllable position and according to the identity of the vowel neighboring the /r/. The neighboring vowel succeeds a prevocalic /r/ and precedes a postvocalic /r/. When neighboring vowels are considered, differences between the formant frequencies in /r/ and in the neighboring vowel can be calculated. This allows a consideration of the formant trajectories between /r/ and its neighboring vowel, which further characterizes the prevo-

TABLE IV. Number of tokens of prevocalic and postvocalic /r/ in vowel context as a function of age category.

Months	Prevocalic /r/ in /ɛ/ context	Prevocalic /r/ in /ɑ/ context	Postvocalic /r/ in /ɛ/ context	Postvocalic /r/ in /ɑ/ context
15–16	0	0	1	0
17–18	0	1	1	0
20–21	1	2	5	4
22–23	1	7	25	11
24–25	4	2	39	19
26–27	5	4	17	16
28–29	4	6	18	18
30–31	8	0	40	6

calic and postvocalic /r/. For example, the differences in $F2$ in /r/ and the $F2$ in the neighboring vowel can be considered. Also, differences in formant frequency separation, i.e., differences in $F3-F2$, for /r/ and for the neighboring vowel can be considered. First, productions of /r/ in prevocalic and postvocalic position in two different vowel contexts are discussed. Then, briefly, the formant frequency changes during syllabic /r/ are examined.

The vowels chosen for examination were the most common in the subjects' productions with prevocalic and postvocalic /r/. These vowels were /ɛ/ and /ɑ/ as in the words "red" and "frog," respectively, or the words "there" and "car," respectively. Table IV shows the number of tokens of each type of /r/ as a function of age category. Focusing on age categories greater than 20–21 months, Fig. 4 reveals that the mean separation between the third and second formant frequencies, $F3-F2$, was actually larger in the prevocalic /r/ than in a succeeding /ɛ/, while this separation was constant in the mean between a postvocalic /r/ and its preceding /ɛ/. Formant separation behavior was different for the back vowel /ɑ/ compared to that of the front vowel /ɛ/. $F3-F2$ was constant in the mean from a prevocalic /r/ to its succeeding /ɑ/, but there was a decrease in mean formant separation

(i.e., smaller $F3-F2$) from an /ɑ/ to its succeeding /r/ (Fig. 4). Also, the behavior of $F2$ depended on whether the neighboring vowel was the front or the back vowel. Figure 5 shows that mean $F2$'s tended to be lower in both the prevocalic and postvocalic /r/ than in the neighboring front vowel /ɛ/. For the back vowel /ɑ/ the mean $F2$'s for prevocalic /r/ were lower than the succeeding /ɑ/, while mean $F2$'s in postvocalic /r/ were higher than in the preceding /ɑ/. These observations of $F2$ help to explain the differences in formant separation, $F3-F2$ behavior in front vowel /ɛ/ context and back vowel /ɑ/ context.

Wilcoxon statistics were used to compare prevocalic /r/ to postvocalic /r/ in /ɛ/ and /ɑ/ contexts. For the /ɛ/ context tokens were from the ages of 28 through 31 months, and for /ɑ/ the tokens were from the ages of 26 through 29 months. (These age categories were chosen to include a sufficient number of tokens to perform statistics, at the same time attempting to limit the range of variation caused by vocal-tract length changes.) Column 1 of Table V shows that the distributions for $F2$ in postvocalic /r/ were stochastically greater than those for prevocalic /r/ with a high probability in both /ɛ/ and /ɑ/ contexts. Similarly, the distributions for $F3$ and $F3-F2$ in postvocalic /r/ were stochastically less than for prevocalic /r/, with a high probability in both vowel contexts (Table V, columns 2 and 3). It can be noted that the probabilities regarding the $F2$ distributions are greater for /ɛ/ than for /ɑ/, and those pertaining to the $F3$ distributions are greater for /ɑ/ than for /ɛ/. This seems to have balanced out to produce very high probabilities for differences between postvocalic and prevocalic /r/ in the distributions for $F3-F2$. Columns 4 and 5 of Table V show that the distributions for the differences between $F3$ and $F3-F2$ in the /r/ and in the vowel were stochastically less than, or stochastically more negative, for postvocalic /r/ than for prevocalic /r/. That is, the *absolute* differences between $F3$ and $F3-F2$
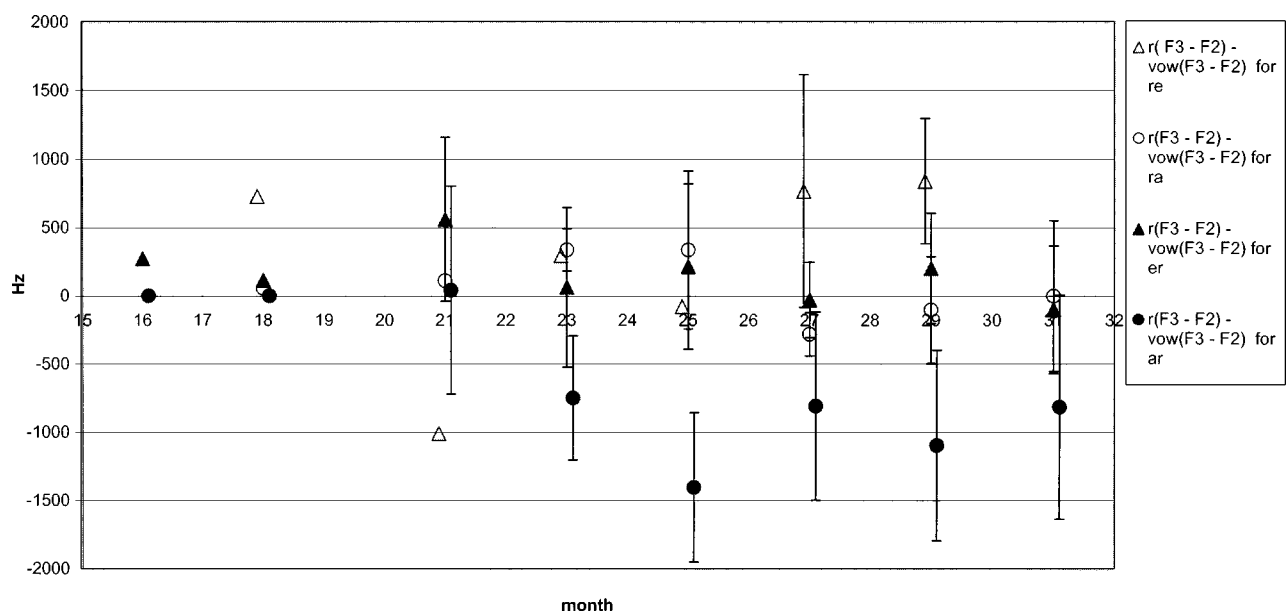

FIG. 4. Means and standard deviations of differences of $F3-F2$ values in the /r/ and the neighboring vowel for prevocalic and postvocalic /r/ in /ɛ/ and /ɑ/ context for the children subjects as a function of age category.
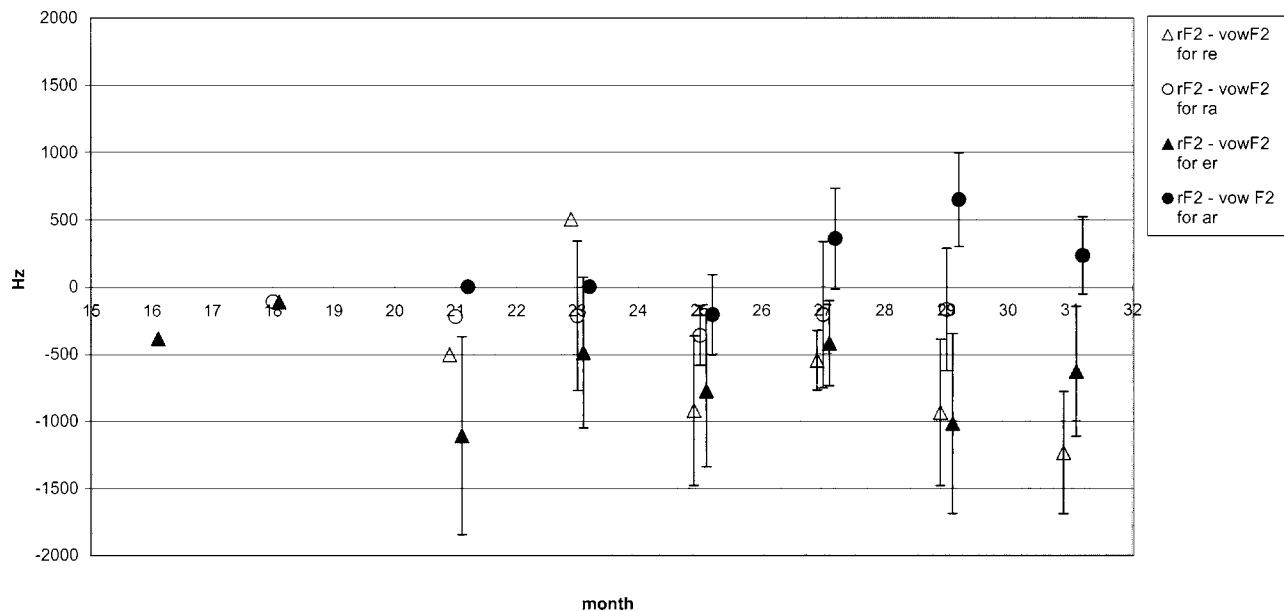
FIG. 5. Means and standard deviations of differences of $F2$ values in the /r/ and the neighboring vowel for prevocalic and postvocalic /r/ in /ɛ/ and /ɑ/ context for the children subjects as a function of age category.

in /r/ and in the vowel tended to be *greater* for postvocalic /r/ than for prevocalic /r/.

The syllabic /r/'s were produced as monophthongs, almost always. Changes in $F2$, $F3$, and $F3-F2$ from the beginning of syllabic /r/ and to the end of each medial and final syllabic /r/ were computed. The mean change in these formant measures was never greater than 500 Hz in magnitude, and most often less than 300 Hz. Further, zero frequency change was within 1 standard deviation of the mean, except in three tokens: (1) one token of $F3-F2$ at 20–21 months for medial syllabic /r/; (2) one token of $F2$ and $F3-F2$ at 14–15 months for final syllabic /r/, and (3) one token of $F3$ at 28–29 months for final syllabic /r/. In the first and third cases, zero frequency change was well within 2 standard deviations of the mean.

IV. DISCUSSION

The results indicate that this group of children was progressing toward postvocalic [ɹ] more rapidly than they were progressing toward prevocalic [ɹ]. The children's $F3-F2$ declined with age faster for postvocalic /r/ than for prevocalic /r/. This result appears to be due to a decrease of $F3$ and a more gradual decrease of $F2$ for postvocalic /r/ with age in comparison with age-related changes in these formant

frequencies for prevocalic /r/ (Table II). Wilcoxon statistics showed that the distributions for $F3-F2$ for final syllabic /r/ and for postvocalic /r/ at 30–31 months were stochastically smaller than the corresponding distributions for prevocalic /r/ with very high probabilities (>0.999) (Table III). While $F3-F2$ of final syllabic /r/ appeared to decrease rapidly with age, there was great variability associated with the slope of the regression line (Table II). On the other hand, $F2$ for final syllabic /r/ increased with age, providing for some of the decrease in $F3-F2$. The behavior of the formant measures for medial syllabic /r/ was more equivocal, probably because the number of tokens of these became relatively small after the 26–27-month category. The functional criterion for including children in the study group (Sec. II A) may have had some effect on the trends in time. However, the comparisons between groups at any age are valid for the particular children at that time.

In a comparison of front and back vowel contexts, /ɛ/ and /ɑ/, it was seen that vowels did not change the statistical relations in the formant frequencies between prevocalic and postvocalic /r/ (Table V). There were, however, effects of vowel context on the $F2$ and formant frequency separations, $F3-F2$ trajectories. For instance, Fig. 5 shows that $F3-F2$ tended to decrease from /ɑ/ into postvocalic /r/, but not for

TABLE V. The probability, to the nearest 0.001: (1) that the $F2$'s for postvocalic /r/'s possess distributions stochastically greater than that for prevocalic /r/'s, and (2) that the $F3$'s and ($F3-F2$)'s for postvocalic /r/ and the differences with their values in the neighboring vowel possess distributions stochastically less than those for prevocalic /r/'s. The age range is 28 to 31 months in the context of vowel /ɛ/ and it is 26 to 29 months in the context of vowel /ɑ/.

Prevocalic versus	$F2$ in /r/	$F3$ in /r/	($F3-F2$) in /r/	$F3$ in /r/- $F3$ in vowel	($F3-F2$) in /r/- ($F3-F2$) in vowel
Postvocalic in /ɛ/ context	0.999	0.995	>0.999	0.994	>0.999
Postvocalic in /ɑ/ context	0.991	>0.999	>0.999	0.996	0.999

/ɛ/ into postvocalic /r/. A rising $F2$ from /ɑ/ into postvocalic /r/ could contribute to a decreasing formant separation, $F3 - F2$. This is consistent with adult formant trajectories for final [ɹ] (e.g., Olive *et al.*, 1993, p. 222), and would indicate a tongue becoming less backed in the transition from the vowel into the liquid. Also, the children seemed to have been producing medial and final syllabic /r/'s as monophthongs, as would be expected in adult production.

The children for whom data are reported were, on the whole, progressing toward postvocalic [ɹ] and final syllabic [ɹ] more rapidly than they were progressing toward prevocalic [ɹ]. While there was acoustic evidence that progress was made by some subjects toward a prevocalic [ɹ], the group as a whole showed little evidence of this. The question naturally arises whether the differences in the acquisition of postvocalic and syllabic versus prevocalic [ɹ] are the result of perceptual or production mechanisms. Further, precisely where is the difficulty? A definite answer to this question cannot be given, but there are some useful pieces of evidence that can be used to argue for certain causes.

In the review of the literature on adult [ɹ] production and in the case of the adult speaker from the XRMB-SPD, prevocalic [ɹ] was identified as having lower $F1$, $F2$, and $F3$ than postvocalic [ɹ], and the prevocalic [ɹ] appears to require two substantial tongue constrictions: one in the oral cavity and the other in the pharynx. The postvocalic [ɹ] and syllabic [ɹ] apparently do not require a tight, or any, pharyngeal constriction. Based on the fact that the newborn infant's tongue fills the oral cavity (Kent and Vorperian, 1995), we would expect that the ratio of the tongue volume to the volume of the supralaryngeal vocal tract is larger for infants and young children than it is for adults. This, together with the fact that the larynx is descending rapidly in the age group considered here (Goldstein, 1980; Kent and Vorperian, 1995), could mean that articulations appropriate for any prevocalic [ɹ] are difficult to attain. Both a bulky tongue body and a small pharyngeal cavity would hinder young children's ability to form both a palatal and pharyngeal constriction with the tongue.

An argument could also be made that the motor control is not mature enough for prevocalic [ɹ] production before, say, 2 or 2 1/2 years. One aspect is that the control of the tongue blade requires some time to mature. Children who are native Spanish speakers master the trilled /r/ at a relatively late age (Jimenez, 1987). We have also noticed that children aged 3 to 7 years make less of an acoustic distinction between /s/ and /ʃ/ than do adults (McGowan and Nittrouer, 1988; Nittrouer, 1995), despite the fact that they can perceive the two phonemes categorically. However, these differences could also be due to the morphology in this region of the tongue and palate. For instance, young children may not be able to create a sufficiently large sublingual cavity for a retroflex [ɹ], just as they have difficulty in producing one for [ʃ] (Perkell, Boyce, and Stevens, 1979; Nittrouer, 1995).

The role of perception in this story is not known, yet its influence cannot be discounted even when there is a plausible articulatory explanation for an observed speech acoustics trend. In fact, there has been some research on the perception and production link of [ɹ] in young children. Menyuk and Anderson (1969) found that their preschool subjects had category boundaries in forced-choice identification between prevocalic /r/, /l/, and /w/. However, according to adult judges, when the children were asked to repeat the words they perceived to be "write" they very often reproduced the word with something close to "white." In general, there was a mismatch between the perceived phoneme and the reproduced phoneme, at least to adult ears, particularly for the /w/–/r/ continuum. The results of the Menyuk and Anderson work indicate some decoupling between perception and production of word initial /r/ in young children. On the other hand, Strange and Broen (1981) found that many of their 3-year-old subjects who tended to produce the word-initial /r/ poorly also tended to be less adult-like in categorization tasks. Thus, the Strange and Broen work emphasized a certain amount of coupling between production and perception capability for word-initial /r/ in children. It should be noted that Strange and Broen used more sophisticated stimuli in simulating their /w/–/r/ continua than did Menyuk and Anderson. The evidence indicates that there is some causal connection between perceptual and production capabilities in learning to produce prevocalic [ɹ], but that one capability does not determine the other. Even for adults who are physiologically capable of producing prevocalic [ɹ] may be able to hear postvocalic [ɹ] more easily than they can hear prevocalic [ɹ], thus leaving them unable to produce prevocalic [ɹ]. This could account for the differences in Japanese identification of [ɹ] and [l] based on syllable position found by Mochizuki (1981).

The previous data of others and current data presented here lead to interesting directions for future research. Below, we briefly discuss some of the speculations and questions raised here. In contrast to prevocalic [ɹ], the young children in this data set were achieving low $F3$ and $F3 - F2$ appropriate for postvocalic [ɹ]. Also, in general, children do not have difficulty producing [w] (e.g., Smits, 1993). Three low formants are required for prevocalic [ɹ], while postvocalic [ɹ] and [w] possess only one or two particularly low formants: $F1$ and $F2$ for [w], and $F3$ for postvocalic [ɹ]. Further, postvocalic [ɹ] and [w] can be articulated using only one oral constriction, other than the lips. The [w] can be articulated using one tongue-body approximation in the velar region and lip rounding. For [w], the first two formants can be modeled using two coupled Helmholtz resonators, and the third formant is the half-wave resonance of either the front or back cavities (Stevens, 1998). Young children should also be able to produce the simple two-cavity system, even if it entails a different cavity affiliation for the third formant than for adults. Similarly, we suggest that postvocalic [ɹ], and syllabic [ɹ] are produced with a single tongue constriction, which is more forward than that for [w]. Particularly, postvocalic [ɹ] behaves as an off-glide to the preceding vowel that young children are able and willing to produce.

Why, if young children can produce postvocalic /r/ with relatively low $F3$ and small $F3 - F2$, don't they do so for prevocalic /r/? In fact many children substitute [w] for prevocalic [ɹ] (Shriberg and Kent, 1995). We speculate that young children tend to substitute [w] for [ɹ] in prevocalic position because they can only produce two low-formant fre-

TABLE VI. The ranges of formant values measured in [ɹ] or syllabic [ɹ] for tokens of several words for JW11 in the XRMB-SPD. The numbers in parentheses are the numbers of tokens analyzed.

	$F1$ range (Hz)	$F2$ range (Hz)	$F3$ range (Hz)	$F4$ range (Hz)	Articulatory properties
"Right" (3)	331 to 441	828 to 961	1255 to 1435	2262 to 2373	Retroflex, lip rounding
"Rag" (2)	386 to 497	938 to 1048	1545 to 1655	2759 to 2814	Nonretroflex, lip rounding tongue distance-to-palate <0.6 cm
"There" (1)	607	1545	1931	2869	Nonretroflex, no lip rounding tongue distance-to-plate >0.8 cm
"Large" (3)	534 to 694	1324 to 1389	1710 to 1843	2814 to 2924	Nonretroflex, no lip rounding tongue distance-to-palate >0.8 cm
"Dormer" (1st syllable) (2)	607 to 662	1102 to 1159	1766 to 1876	2759	Nonretroflex, no lip rounding tongue distance-to-palate >0.8 cm
"Dormer" (2nd syllable) (2)	607	1159 to 1214	1655	2704	Nonretroflex, no lip rounding tongue distance-to-palate <0.4 cm

quencies simultaneously. A reasonable compromise would be achieved in this substitution, because prevocalic [ɹ] possesses relatively low $F1$ and $F2$, and these formants are, perhaps, perceptually more salient than $F3$. Some children appear to make this compromise in articulatory behavior and continue it even after they are physically more capable of producing a prevocalic [ɹ]. The reason that prevocalic [ɹ] is sometimes not forthcoming, even when their productive lexicons requires an /r/–/w/ distinction, is a topic for further research. Perhaps part of the answer is that some of these children make a subphonemic distinction between prevocalic /w/ and /r/, which is an acceptable categorical distinction for them (Hoffman, Stager, and Daniloff, 1983). Also, children who do not start toward an adult-like prevocalic [ɹ] at an early enough age may have difficulty in changing a strong coupling between learned motor behavior and perceptual attention.

V. CONCLUSION

The formant frequency data on this group of children, from about 14 months through 26 months, and some through 31 months, indicated that they were acquiring aspects of [ɹ] production for postvocalic and syllabic /r/ before they were acquiring equivalent aspects of production for prevocalic /r/. That is, the distributions of $F3$'s and $(F3 - F2)$'s for postvocalic, and least for final syllabic, /r/ were stochastically less than those for prevocalic /r/ in the final months of recording. Also, the rate of decline for $F3 - F2$ with age was significantly greater in postvocalic /r/ than in prevocalic /r/. However, the prevocalic /r/ $F2$ distributions were stochastically smaller than those for postvocalic and syllabic /r/'s with a very high probability, which is also typical of the differences seen in adults according to syllable position. A comparison of a front and a back vowel context did not reveal that vowel context affected these basic results. We have attributed part of the reason for lack of progress in prevocalic /r/ production to the complexity of its articulation. However, this is not the complete explanation and more research into the relation between perception and production needs to be pursued for a complete picture.

APPENDIX: A CASE STUDY ON AN ADULT

The subject JW11 from the XRMB-SPD was examined in some of his productions of [ɹ], because he used both retroflex and nonretroflex articulations to produce prevocalic [ɹ]. Table VI presents the ranges of the first four formant frequencies and their articulatory correlates for JW11 in [ɹ] for each of the words "right," "rag," "there," "large," and "dormer."

There was a tendency for all formant frequencies to be lower for the retroflex prevocalic [ɹ] than for nonretroflex prevocalic [ɹ] for all formants, as shown in rows 1 and 2 of Table VI. This was particularly clear for $F4$, and this corroborates the 1999 finding of Espy-Wilson and Boyce.

Each postvocalic [ɹ] was produced with a nonretroflex articulation, and formant frequencies, are shown in rows 3 through 5 of Table VI. The first three formant frequencies for all postvocalic [ɹ]'s were higher than the corresponding formant frequencies for both the retroflex and nonretroflex articulations of prevocalic [ɹ]. In particular, even with the relatively low $F2$, caused by the backing of the tongue for the preceding vowel in "dormer," the $F2$'s for postvocalic [ɹ]'s were all greater than for the prevocalic [ɹ]'s. The $F4$'s for postvocalic [ɹ] were similar to those of the nonretroflex prevocalic [ɹ].

The second syllable of the two tokens of "dormer" were examined as examples of syllabic [ɹ] in row 6 of Table VI. These were produced as nonretroflex articulations and only slightly diphthongized. Except for the $F3$'s as low as the highest value for a prevocalic [ɹ] production, the formant frequency ranges were in similar relation to those of the prevocalic [ɹ]'s as were those of the postvocalic [ɹ]'s.

Pellet positions were examined to discover the articulatory bases for the measured formant frequencies. JW11 produced lip rounding during prevocalic [ɹ], but not for postvocalic or syllabic [ɹ]. This helped to account for at least some of the differences in formant frequencies between nonretroflexed prevocalic [ɹ] and postvocalic [ɹ].

In consideration of articulatory factors, the nonretroflexed [ɹ] is examined first. The two chosen examples of the word "rag" spoken by JW11 possessed nonretroflex [ɹ]. The constriction for [ɹ] was forward of the constriction for [g] (i.e., it was more palatal than velar). However, the tongue was more "bunched" for [ɹ] compared to [g], so that the tongue tip was further back and higher for [ɹ] than for [g]. Also, lip rounding occurred for prevocalic [ɹ], but not for [g]. This caused the $F3$ for [ɹ] to be lower than for [g] because the front cavity served as a Helmholtz resonator with resonance frequency $F3$ in [ɹ] whose volume (capacitance) and mass elements are maximized. (The $F3$ [g] was measured during the burst.) On the other hand, for [g] in a front vowel context, $F3$ can be approximately a quarter-wave resonance of the front cavity. The predictions for $F2$ in these segments are more difficult because there is no indication of the tongue root or larynx position in the XRMB-SPD data set. In fact, $F2$ for [ɹ] was substantially lower (938 and 1048 Hz) than it was for [g] (1644 and 1821 Hz), which is consistent with the palatal constriction being more forward and the tongue more bunched for [ɹ] than for [g].

What was the difference in JW11's articulation between nonretroflex prevocalic [ɹ], and postvocalic and syllabic [ɹ], other than lip rounding? An examination of the articulatory data reveals that the difference was partly one of palatal constriction degree, as measured by the minimum distance of the pellets to the palate. The palatal approximation that the tongue made was not as narrow for the postvocalic [ɹ] in "there," "large," and "dormer" as it is for the nonretroflexed prevocalic [ɹ] in "rag." In the midsagittal plane, postvocalic [ɹ] had a minimum postalveolar pellet-to-palate distance of at least 0.8 cm, while for prevocalic [ɹ] this distance was less than 0.6 cm. A lower tongue position for postvocalic [ɹ] compared to nonretroflex prevocalic [ɹ] is consistent with the higher $F1$ in postvocalic [ɹ] observed for JW11. The higher $F3$ in postvocalic [ɹ] compared to prevocalic [ɹ] is consistent with reduced lip rounding and palatal constriction degree in the former compared to the latter. Perhaps because the syllabic [ɹ] tokens from "dormer" tokens were the final syllable in a word from a word list, these were more strongly articulated than they would be in other situations. The $F3$'s from the syllabic [ɹ]'s were low compared to those of the postvocalic examples from JW11. Views of JW11's articulations reveal that these syllabic [ɹ]'s are produced with tight postalveolar or palatal constrictions (pellet-to-palate distances were less than 0.4 cm). The moderate $F1$'s may have been the result of a reduced palatal constriction length that decreased the effective mass of the Helmholtz resonator that gave rise to $F1$.

[1] The various phonetic variants of American English /r/ will be denoted collectively by the symbol [ɹ]. The phoneme symbol /r/ will be employed when discussing children's productions of words that would contain [ɹ] in an adult's production, rather than referring to "attempted productions of [ɹ]."

[2] The spacing between adjacent formants should be at least 1400 Hz for 4-year-old children, based on data provided by Kent and Forner (1979). This estimate is based on a formant scale factor of at least 40% and an average adult male frequency spacing of 1000 Hz. The formant spacing should be even greater for children less than 2.5 years old. The same data indicate that the $F3$ for neutral vowels is greater than 3500 Hz for the group of children considered here.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). "Toward articulatory–acoustic models for liquid approximants based on MRI and EPG data. II. The rhotics," J. Acoust. Soc. Am. **101**, 1078–1089.

Bickel, P. J., and Doksum, K. A. (**1977**). *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden-Day, San Francisco).

Dalston, R. (**1975**). "Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults," J. Acoust. Soc. Am. **57**, 462–469.

Delattre, P., and Freeman, D. C. (**1968**). "A dialect study of American r's by X-ray motion picture," Linguistics **44**, 29–68.

Espy-Wilson, C. Y., and Boyce, S. (**1999**). "The relevance of $F4$ in distinguishing between different articulatory configurations of American English," J. Acoust. Soc. Am. **105**, 1400.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., and Alwan, A. (**2000**). "Acoustic modeling of American English /r/," J. Acoust. Soc. Am. **108**, 343–356.

Fisz, M. (**1963**). *Probability Theory and Mathematical Statistics*, 3rd ed. (Wiley, New York).

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, MIT, Cambridge, MA).

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. S. (**1999**). "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," J. Acoust. Soc. Am. **105**, 2854–2865.

Hagiwara, R. (**1994**). "Three types of American /r/," UCLA Working Papers of Phonetics, **88**, 55–61.

Hoffman, P. R., Schuckers, G. H., and Ratusnik, D. L. (**1977**). "Contextual-coarticulatory inconsistency of /r/ misarticulation," J. Speech Hear. Res. **20**, 631–643.

Hoffman, P. R., Stager, S., and Daniloff, R. G. (**1983**). "Perception and production of misarticulated /r/," J. Speech Hear Disord. **48**, 210–248.

Jackson, M. T.-T., Espy-Wilson, C. Y., and Boyce, S. E. (**2001**). "Verifying a vocal tract model with a closed side-branch," J. Acoust. Soc. Am. **109**, 2983–2987.

Jimenez, B. C. (**1987**). "Acquisition of Spanish consonants in children aged 3–5,7," Lang. Speech Hear. Serv. Schools **18**, 357–363.

Kent, R. D., and Forner, L. L. (**1979**). "Developmental study of vowel formant frequencies in an imitation task," J. Acoust. Soc. Am. **65**, 208–217.

Kent, R. D., and Vorperian, H. K. (**1995**). "Development of the craniofacial–oral–laryngeal anatomy: A review," J. Medical Speech-Language Pathology **3**, 145–190.

Lehiste, I. (**1962**). *Acoustical Characteristics of Selected English Consonants* (Communication Sciences Laboratory, University of Michigan, Report No. 9, Ann Arbor, MI).

McGowan, R. S., and Nittrouer, S. (**1988**). "Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/," J. Acoust. Soc. Am. **83**, 229–236.

Menyuk, P., and Anderson, S. (**1969**). "Children's identification and reproductions of /w/, /r/, and /l/," J. Speech Hear. Res. **12**, 39–52.

Mochizuki, M. (**1981**). "The identification of /r/ and /l/ in natural and synthesized speech," J. Phonetics **9**, 283–303.

Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**, 520–530.

Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech* (Springer, New York).

Perkell, J. S., Boyce, S. E., and Stevens, K. N. (**1979**). "Articulatory and acoustic correlates of the [s–ʃ] distinction," J. Acoust. Soc. Am. **65**, S24.

Sander, E. K. (**1972**). "When are speech sounds learned?" J. Speech Hear Disord. **37**, 55–63.

Shriberg, L., and Kent, R. (**1995**). *Clinical Phonetics* (Allyn and Bacon, Needham Heights, MA).

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., and Bird, A. (**1990**).

"The Iowa articulation norms and its Nebraska replication," J. Speech Hear Disord. **55**, 779–798.

Smit, A. B. (**1993**). "Phonological error distributions in the Iowa–Nebraska articulation norms project: Consonant singletons," J. Speech Hear. Res. **36**, 533–547.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stoel-Gammon, C. (**1985**). "Phonetic inventories, 15–24 months: A longitudinal study," J. Speech Hear. Res. **28**, 505–512.

Strange, W., and Broen, P. A. (**1981**). "The relationship between perception and production of /w/, /r/, and /l/ by three-year-old children," J. Exp. Child Psychol. **31**, 81–102.

Westbury, J. R. (**1994**). "X-ray Microbeam Speech Production Database User's Handbook" (University of Wisconsin, Madison, WI, unpublished).

Westbury, J. R., Hashi, M., and Lindstrom, M. J. (**1998**). "Differences among speakers in lingual articulation for American English /ɹ/," Speech Commun. **26**, 203–226.