JSLHR

Research Article

A Longitudinal Study of Very Young Children's Vowel Production

Rebecca W. McGowan,[a] Richard S. McGowan,[a] Margaret Denny,[a] and Susan Nittrouer[b]

Purpose: Ecologically realistic, spontaneous, adult-directed, longitudinal speech data of young children were described by acoustic analyses.
Method: The first 2 formant frequencies of vowels produced by 6 children from different American English dialect regions were analyzed from ages 18 to 48 months. The vowels were from largely conversational contexts and were classified according to dictionary pronunciation.
Results: Within-subject formant frequency variability remained relatively constant for the span of ages studied. It was often difficult to detect overall decreases in the first 2 formant frequencies between ages 30 and 48 months. A study of the movement of the corner vowels with respect to the vowel centroid showed that the shape of the vowel space remained qualitatively constant from 30 through 48 months.
Conclusions: The shape of the vowel space is established early in life. Some aspects of regional dialect were observed in some of the subjects at 42 months of age. The present study adds to the existing data on the development of vowel spaces by describing ecologically realistic speech.

Key Words: children, development, speech production, acoustics

Speech acquisition by children involves all aspects of development, including motor control, perception, cognition, and the physical growth of body structures. As children acquire the ability to speak in a language environment, they often speak to be understood by adults who possess a working knowledge of that language. The degree to which children who are just learning to speak have an understanding of their language's structure, including its phonology, has not been established. Thus, an important avenue of research is to describe the speech production of children who are learning to speak their ambient language concurrent with gaining a working knowledge of its structure. One way to describe the process of learning vowel production is to measure the acoustic parameters of vowels as they are produced in meaningful utterances directed toward adults.

The present study is a descriptive one, based on acoustic speech data collected for other purposes (Nittrouer, 2010). The first two formant frequencies, F1 and F2, were used to follow the acoustic phonetic development of the monophthongal vowel production of six American children from age 18 to 48 months as they learned to speak English. The speech samples were recorded at 6-month intervals in the company of English-speaking adults. The vowels were classified according to their dictionary phonemic identity in the words spoken by the children, and therefore the so-called Standard American English pronunciations are identified. In this way, the acoustic phonetic aspect during the acquisition of the vowel portion of the children's phonological system can be described and documented.

Ecologically realistic, spontaneous, adult-directed speech—from conversation and from elicitation by pictures—and a small number of words elicited by imitation with vowels in various phonetic contexts were included in the present longitudinal study. Formant measurement is performed on words that could be understood by the authors in the context in which they were spoken. The following review of the literature shows that the kinds of data examined in the present study are unique. The methods by which the data were collected are the focus, and the results of many of these studies are noted later in the Results section of the present study.

Several studies have measured vowel formant frequencies from speech produced by children and adults to understand speech development (see Vorperian & Kent, 2007, for a review). Peterson and Barney (1952) measured formant frequencies of vowels spoken by children. The ages and place of residence of the children were not specified, but the children must have been old enough to read from the elicitation lists of /hVd/ words. The classic study by Eguchi and Hirsh

[a]CReSS LLC, Lexington, MA
[b]The Ohio State University, Columbus
Correspondence to Margaret Denny: megdenny@comcast.net

Editor and Associate Editor: Jody Kreiman
Received April 6, 2012
Revision received September 5, 2012
Accepted April 21, 2013
DOI: 10.1044/1092-4388(2013/12-0112)

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

(1969) included preliterate and literate children and measured F1 and F2 of vowels in two short sentences spoken by children ages 3 to 13 years and adults from the St. Louis, Missouri, area. Each vowel of interest was contained once in either of these sentences. For children younger than age 7 years, the sentences were elicited by imitation; older children and adults read the sentences aloud.

As formant tracking technology has improved, it has been possible to consider large groups of children and to track the trajectories of their vowel formants automatically. Hillenbrand, Getty, Clark, and Wheeler (1995) extended Peterson and Barney's (1952) study of vowels by analyzing the speech of 46 ten- to twelve-year-old children and 93 adults primarily from Michigan. Speakers read lists of /hVd/ words. Looking more broadly at development, Lee, Potamianos, and Narayanan (1999) studied changes in formant frequencies, among other acoustic measures, in the speech of 436 children ages 5 to 18 years and in the speech of adults. The subjects, who were primarily from Missouri and Illinois, were recorded speaking target /bVt/ words in a carrier sentence as well as five meaningful sentences. The speech tokens were elicited with written words and sentences, except in the case of some 5- to 7-year-olds, who required elicitation by imitation. Assmann and Katz (2000) made formant measurements for children at ages 3, 5, and 7 years and for adults from Dallas, Texas, for a perceptual study. Because of their study's focus, a subset of the vowels was analyzed on the basis of "adequate pronunciation quality." The vowels were spoken in /hVd/ contexts in imitation of a recording of an adult speaker. As part of a larger study, Ménard, Schwartz, Boë, and Aubin (2007) recorded 4-year-olds, 8-year-olds, and adults producing Canadian French vowels (five each) elicited in a carrier phrase after an adult speaker produced the phrase.

Common to all of the studies reviewed above is the fact that word elicitation was achieved through either reading or imitation, depending on the age of the subject. Some studies used only reading, whereas others required a mix of these elicitation methods. Some studies have used only elicitation by imitation to study the vowel production of preliterate 4-year-old children (see, e.g., Kent & Forner, 1979). The phonetic contexts of the vowels of interest were fixed in each of these studies.

Obtaining and processing speech samples of children younger than age 3 years is a major difficulty. High and highly variable fundamental frequencies (F0s) make automatic formant tracking unreliable. For very young children (see, e.g., under age 15 months), it is common to study acoustics of vowel-like spontaneous productions without reference to category, as Kent and colleagues have done (see, e.g., Kent & Murray, 1982; Kent, Osberger, Netsell, & Hustedde, 1987; McGowan, Nittrouer, & Chenausky, 2008). In studies of slightly older children, categories based on acoustic features have been used for the study of spontaneous speech. For example, Gilbert, Robb, and Chen (1997) measured formant frequencies in children over the course of ages 15 to 36 months and categorized vowels by tongue height and advancement.

In the present longitudinal study, we used a method that differs from those used in previous studies. Six children from ages 18 months to 48 months were followed longitudinally in their speech production. This study is unique in that we examined largely spontaneous speech whereby the vowels are grouped according to dictionary phonemic classification. There are various ways to consider the data descriptively. With dictionary classification, the growth of the shape of individual vowel spaces, including aspects of dialect, can be documented as children accumulate their productive vocabularies. Issues of variability are always important when studying the development of children's speech. Later in this article, we make comparisons between the present research and previous studies that have used other elicitation methods. Although there are advantages to the methods we used in the present study, this method will be unable to follow development of fine phonetic detail because of the many factors that cannot be controlled when spontaneous speech is the main source of data. The previous studies are advantageous in this respect.

In the present article, we explore topics in a descriptive manner. We consider how our results differ from those of many of the previous studies described above. We compare the studies in terms of variability and changes of mean formant frequencies with age. We expected that the vowels in the present study would exhibit greater variability than those in previous studies because we used spontaneous speech data. In this study, we also examined indications of the development of regional dialect. The question as to when the basic shape of the vowel space is established is important and was also considered. We were unable to explore effects such as gender on vowel production because of the relatively small subject pool (Perry, Ohde, & Ashmead, 2001).

Method

Subjects

The six children studied in the present work were American English speakers from three dialect areas: Northern (one from Rochester, Minnesota, and two from Chicago, Illinois), Midland (one from Oklahoma City, Oklahoma), and West (two from Logan, Utah) (Labov, Ash, & Boberg, 2006). The recordings were originally made in a study investigating language development in children with and without hearing loss (Nittrouer, 2010). The speakers all passed universal newborn hearing screenings and, at age 36 months, passed hearing screenings consisting of the pure tones 0.5, 1.0, 2.0, and 4.0 kHz presented at 20 dB HL to each ear separately.

Materials

The children were recorded at 6-month intervals from ages 18 to 48 months as they played with toys with a parent (spontaneous conversational speech) as well as named vocabulary items from pictures (expressive vocabulary samples) and repeated words after an experimenter (isolated target words). The last types of recordings were made only at ages 36, 42, and 48 months as the children's speech intelligibility was tested using Wilcox and Morris's (1999) Children's Speech Intelligibility Test. The play recordings were made at all ages and were approximately 20 min long. A fourth type of recording was made of the mother giving direction to the child; the child may often be heard clearly, as well, in these recordings. These recordings were approximately 10 min long and supplemented the above materials. The percentage of vowels in this study that are from conversational speech are, for each subject, in increasing order, 73%, 73%, 80%, 88%, 93%, and 94%.

Each subject's speech was transcribed phonetically and orthographically. We were interested in learning about how children produced vowels that were directed toward adults. Vowels found in words that we understood were categorized according to what is referred to in the present work as the *dictionary vowel,* rather than by their pronunciation. Dictionary vowels were determined by the standard pronunciation given in the *American Heritage College Dictionary* (1993). Stressed vowels were easily categorized, unless the dictionary gave multiple accepted pronunciations, in which case we did not analyze the vowel. We typically did not measure vowels produced without stress, unless the dictionary pronunciation was /ʌ/ (i.e., we did not measure vowels that the dictionary categorized as /ə/ or those that the dictionary said were nonreduced vowels if the production was highly reduced).

This approach to categorizing target vowels does not take the dialect's target vowel into account. Thus, when formant frequencies are used to characterize these orthographically transcribed vowels, the effects of dialect can become apparent. For example, perhaps a dictionary /u/ is more fronted for the Western speakers than standard pronunciation would dictate and production of /u/ in these speakers may not relate to their ability to produce a backed vowel.

We did not analyze vowels found in a nasal context (e.g., *mom*) because of the potential for nasal formants; neither did we analyze vowels that were intended to be followed by /ɹ/ (e.g., *our*) due to the possibility of *r*-coloring of the vowel. Stressed vowels following /ɹ/ were measured. If a vowel had a very high F0 (with F0 higher than the expected F1 frequency region), we did not analyze it because of the difficulty in making accurate formant measures.

Measurements

The sound files, recorded at 48 kHz, were decimated to a sampling rate of 16 kHz. F1 and F2 were measured by hand from the cross-sectional spectrum while simultaneously viewing the spectrogram using the speech analysis software Speech Station II and Wave Surfer. (Two programs were used because operating system changes made the former program unusable.) Spectra were plotted from a discrete Fourier transform analysis with a Hamming window size of 512 samples without preemphasis.

In this article we present only the results from the vowels /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, ʊ, u/ that were produced as monophthongs. The vowels /e/ and /o/ were not included because they are typically diphthongized in American English. Vowels with steady formants in the middle of the vowel were measured once in the middle 50% of the vowel. A formant frequency was estimated by interpolating, by eye, among the three highest voice harmonics in the vicinity of the formant in the spectral cross-section. The frequency vicinities of F1 and F2 were estimated from the spectrogram. We also attempted to measure F3 in order to ensure that the measured F2 was not confused with F3. The F3 measurements were not used in the present study because they could not always be reliably measured. If the vowel was diphthongized with formant movement visible on the spectrogram, it was not analyzed. An exception to the method described above was made for /u/, which was often highly yodded in some contexts, such as /d/. When the formant frequencies decreased markedly over the course of /u/, just one measurement was made toward the end of the vowel.

Measurements on each subject were made by one author, and another author made measurements on 10% of these vowels to verify the reliability of the F1 and F2 measurements. The resulting reliability measures are shown in Table 1 in terms of the percentage absolute difference in the two measurements. The average percentage difference in F1 (6.9%–9.7%) is about twice that for F2 (3.6%–4.6%). We deemed these reliability measures acceptable.

Results

Frequency of Vowels

The frequency of vowels measured is not representative of the frequency of production by the subjects, because the analyzed vowels are those within the words that were understood and not excluded for other reasons (nasal context, rhotacization). In general, it is true that there are more tokens at the later ages due to a combination of the children speaking more and being more intelligible. The number of tokens for each vowel is shown in the Appendix.

In addition to determining the dictionary vowel, which was used as the basis of analysis in this study, each vowel was also transcribed phonetically. Actual vowel production is compared to dictionary vowels in Table 2. We discuss this, along with observations on dialect features, below. The two transcribers were originally from suburban Boston, Massachusetts, and central New Jersey, and both were living in the Boston area. Recall that vowels for which the production was reduced or diphthongized are not included in this analysis. The dictionary vowel was different from the transcribed vowel for 8.0% of dictionary /æ/ tokens, 9.1% of dictionary /i/ tokens, 9.9% of dictionary /ʌ/ tokens, 11.5% of dictionary /u/ tokens, 11.8% of dictionary /ɪ/ tokens, 14.3% of dictionary /ʊ/ tokens, 14.6% of dictionary /ɑ/ tokens, 14.7% of dictionary /ɛ/ tokens, and 40.6% of /ɔ/ tokens. In terms of transcription, the vowels at the corners of the vowel space were produced more accurately than were the others. With the exception of /ɛ/, the front vowels were produced more accurately than the back vowels. The low-back vowels were among the least accurately produced vowels.

Formant Frequencies

Vowel spaces at each age are plotted in Figure 1. Mean formant frequencies at each age are given in Table 3. Each value shown is the grand mean of the subjects' means of all

Table 1. Reliability of the formant frequency measures as a percent of absolute difference.

Subject formant	Rochester, MN		Chicago, IL 1		Chicago, IL 2		Oklahoma City, OK		Logan, UT 1		Logan, UT 2	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
M	6.9%	4.6%	7.8%	4.0%	7.8%	4.5%	9.7%	4.5%	8.0%	4.5%	8.2%	3.5%
SD	5.4%	5.4%	6.6%	4.2%	6.8%	4.4%	9.5%	6.7%	5.2%	3.8%	7.1%	2.9%
n	51	51	158	158	117	117	93	93	68	68	136	136

tokens. The standard deviation, shown in parentheses, is the dispersion of the subjects' means about the grand mean—that is, the between-subject variability.

Direct comparison is possible between the formant frequency values measured in the present study and those used by Eguchi and Hirsh (1969), who measured the formant frequencies in the vowels /i, ɛ, æ, ɔ, u/ of five children at ages 36 and 48 months. Assmann and Katz (2000) measured formant frequencies in vowels of three children at age 36 months. The results from these two studies are plotted with the current results in Figure 2. Bars indicate the standard deviation between subjects in each study.

For mid- and low vowels, F1 values were generally lower in Eguchi and Hirsh's (1969) data than in the present data. Most notable are a low F1 in /æ/ and the resulting similarity between /æ/ and /ɛ/ in Eguchi and Hirsh's data as compared to the present study at both 36 and 48 months. Otherwise, F1 and F2 values are generally quite similar between the present study and Eguchi and Hirsh's data as well as between the present study and Assmann and Katz's (2000) data.

Formant Frequencies Over Time

To examine the relation of formant frequencies to time, we computed hierarchical multiple regressions with the dependent variables F1 or F2 and the factors vowel, and then age, for each subject using SPSS. The first factor, vowel, was expected to account for a large amount of variation in F1 and F2 because the vowels are distinct in F1–F2 space; adding age determines how much additional variation is accounted for by the subject's age. Hierarchical regression can be effective when the first group of factors, denoted as *control variables,* is known to have an effect on the dependent variable, and the subsequent group of factors, denoted *predictor variables,* is the set on which interest is focused. Here, the vowel identity is the control variable, and subject age is the predictor variable. Thus, for each subject and formant frequency, the first regression was a one-way analysis of variance with Vowel Identity as a factor. The second regression added age as a factor, which is similar to an analysis of covariance, but instead of partialing out the effect of the covariate age, the factor of Vowel Identity is partialed out to find the effect of age on formant frequency (Bickel & Doksum, 1977; Cohen, Cohen, West, & Aiken, 2003; Field, 2009).

Table 2. Dictionary vowel (across top) compared to vowel observed (along left).

Number of vowels observed	i	ɪ	ɛ	æ	ʌ	ɑ	ɔ	ʊ	u
i	914	26	3	1	0	0	0	0	3
ɪ	80	1,005	15	6	6	0	0	2	3
ɛ	2	88	320	16	4	1	0	2	0
æ	2	2	10	567	0	24	4	0	0
ʌ	1	8	12	5	366	4	4	15	5
ɑ	1	3	13	19	17	197	38	1	0
ɔ	0	0	1	0	1	4	76	1	2
ʊ	0	6	0	0	10	0	0	162	26
u	0	1	0	0	2	0	0	4	374
e	7	1	0	1	0	0	0	0	3
o	0	0	1	1	0	1	6	2	2

The vowel categories were coded with unweighted effects coding. This makes each vowel independent variable a comparison between the vowel and the centroid (the unweighted mean of vowels)—specifically, the difference between the mean formant frequency for the vowel and the centroid for that age were regressed onto age (Cohen et al., 2003). For reasons we discuss in the Shapes and Positions of Vowel Spaces section, the centroid was determined to be useful for ages 30 months and older in the present data; thus, the regressions were computed for ages 30 through 48 months.

The regression analyses were used as descriptive statistics for the present data set and should not be viewed as predictive statistical models (Hays, 1981, pp. 459–461). Vowel alone accounted for a substantial amount of F1 variation for each subject (between 60.7% and 69.8%, $p < .001$). Furthermore, age accounted for a certain amount of additional variation of F1 in three of the subjects (between 0.9% and 5.9% additional variation, $p < .001$). These three subjects who showed statistically significant F1 decreases with age were Chicago, IL 1 (slope = –3.57 Hz/month); Chicago, IL 2 (slope = –4.57 Hz/month); and Rochester, MN (slope = –9.16 Hz/month). The effects of the addition of age to the model are moderate in two further subjects: Logan, UT 2 ($p = .003$ and slope = –2.18 Hz/month) and Logan, UT 1 ($p = .005$ and slope = –3.45 Hz/month).

Vowel accounts for a large amount of the variation in F2 for each subject (between 73% and 81.7%, $p < .001$). The addition of age into the model accounts for a small amount of additional variation in two of the subjects (between 0.6% and 0.7% additional variation, $p < .001$) These subjects were from the Chicago, IL 1 (slope = –8.04 Hz/month) and Logan, UT 2 (slope = –9.23 Hz/month) groups.

Within-Subject Formant Frequency Variability

Studies of children's vowel formant frequency development have often concluded that children achieve vowel

Figure 1. Vowel spaces from grand mean formant frequencies at 18 months (Panel A), 24 months (Panel B), 30 months (Panel C), 36 months (Panel D), 42 months (Panel E), and 48 months (Panel F). Mean formant frequencies for each vowel are plotted with 1-*SD* bars in the F1 and F2 directions. Error bars indicate the standard deviation between subjects in each study.

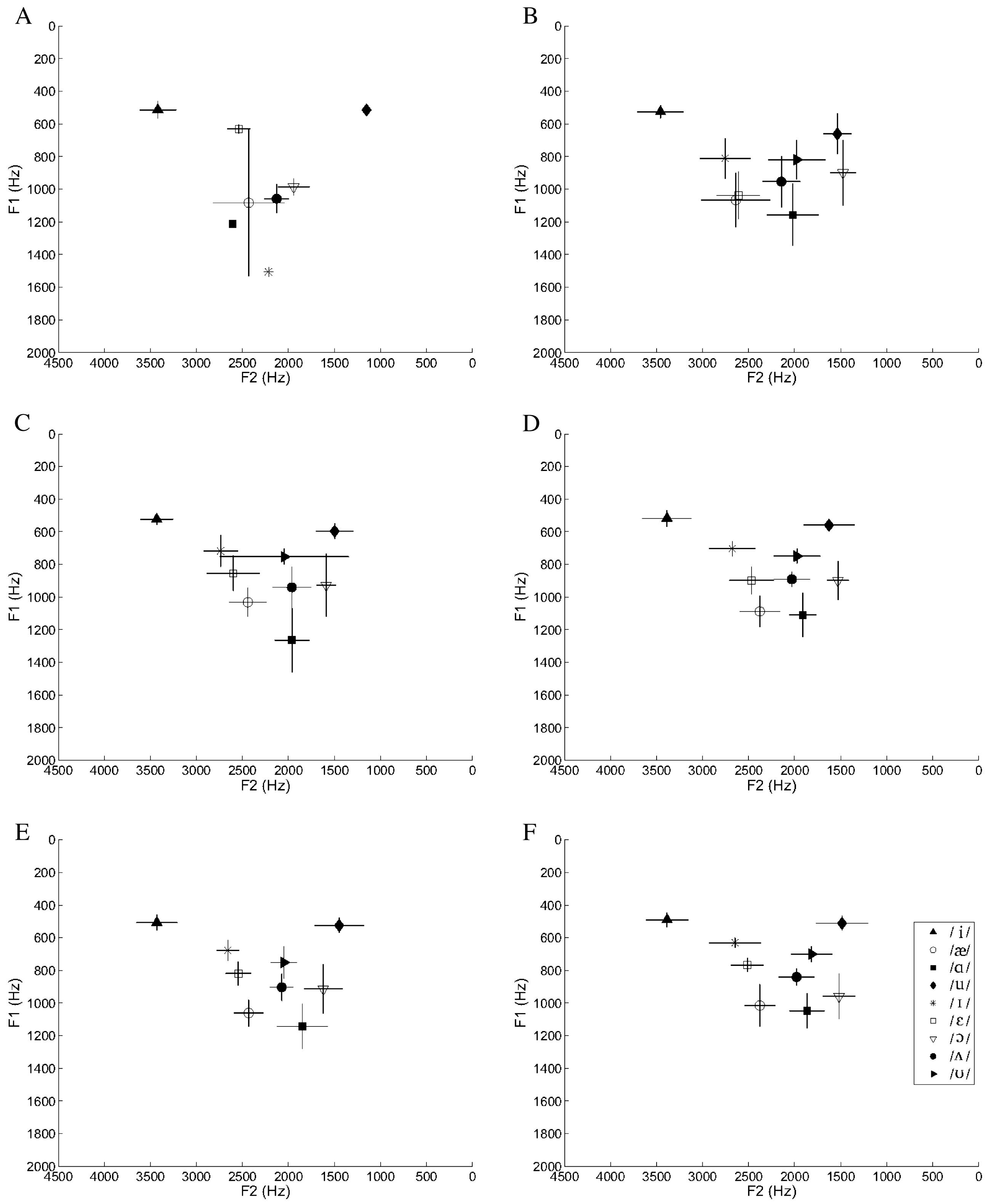


Table 3. Formant frequencies at all ages: *M*s and *SD*s.

		Age (months)					
Vowel	Variable	18	24	30	36	42	48
i	*n*	6	6	6	6	6	6
	F1	514 (51)	527 (38)	523 (31)	518 (49)	508 (45)	491 (43)
	F2	3,418 (197)	3,458 (246)	3,430 (175)	3,386 (265)	3,429 (221)	3,386 (226)
ɪ	*n*	1	6	6	6	6	6
	F1	1,506 (—)	811 (122)	718 (94)	704 (45)	678 (60)	631 (29)
	F2	2,214 (—)	2,753 (271)	2,735 (183)	2,678 (247)	2,656 (118)	2,648 (279)
ɛ	*n*	2	6	6	6	6	6
	F1	631 (26)	1,038 (143)	854 (106)	898 (83)	819 (71)	767 (40)
	F2	2,539 (125)	2,610 (230)	2,598 (284)	2,468 (238)	2,544 (135)	2,513 (174)
æ	*n*	3	6	6	6	6	6
	F1	1,083 (448)	1,066 (165)	1,032 (88)	1,088 (93)	1,062 (79)	1,015 (128)
	F2	2,430 (384)	2,638 (370)	2,440 (201)	2,378 (216)	2,431 (155)	2,378 (164)
ʌ	*n*	3	4	6	6	6	6
	F1	1,057 (85)	954 (155)	940 (125)	890 (44)	904 (81)	841 (52)
	F2	2,127 (131)	2,142 (203)	1,962 (210)	2,029 (190)	2,071 (124)	1,977 (189)
ɑ	*n*	1	4	5	6	6	6
	F1	1,211 (—)	1,157 (189)	1,264 (196)	1,110 (132)	1,143 (135)	1,049 (106)
	F2	2,607 (—)	2,020 (278)	1,960 (187)	1,911 (143)	1,848 (276)	1,865 (187)
ɔ	*n*	2	2	2	5	6	6
	F1	986 (49)	899 (200)	927 (192)	898 (117)	914 (149)	959 (138)
	F2	1,942 (170)	1,474 (136)	1,588 (101)	1,530 (115)	1,620 (206)	1,517 (176)
ʊ	*n*	0	3	4	5	6	6
	F1	—	819 (118)	753 (46)	750 (43)	751 (98)	701 (47)
	F2	—	1,976 (309)	2,044 (697)	1,974 (250)	2,050 (138)	1,815 (222)
u	*n*	1	4	6	6	6	6
	F1	515 (—)	661 (122)	597 (47)	560 (34)	525 (45)	512 (42)
	F2	1,151 (—)	1,534 (149)	1,496 (201)	1,626 (275)	1,446 (266)	1,484 (280)

Note. *SD*s are shown in parentheses; they represent between-subject variability. *n* = number of subjects with tokens. Dashes indicate instances in which a *M* or *SD* could not be computed.

targets more accurately as they get older because vowel formant frequencies become less variable with age (Assmann & Katz, 2000; Eguchi & Hirsh, 1969; Lee et al., 1999; Ménard et al., 2007). A decrease in acoustic variability is often interpreted as an increase in articulatory precision (Vorperian & Kent, 2007) resulting from the maturation of the motor control system (Smith, 2010). There is also evidence that children's speech is more strongly influenced by intrasyllabic coarticulatory effects than is the speech of adults (Nittrouer, Studdert-Kennedy, & McGowan, 1989; Nittrouer, Studdert-Kennedy, & Neely, 1996), which could also influence the measured variability of vowel formant frequencies. In this study, we examined variability over time among vowels found in the varied phonetic contexts of running, natural speech.

We calculated a measure of F1 or F2 within-subject variability that summarizes all subjects following Eguchi and Hirsh (1969). For each vowel, the square root of the mean of individual subjects' variances was divided by the grand mean frequency across subjects for that vowel. This is known as *normalized within-subject variability* and represents typical variability. The results are shown in Figure 3, in which the variability in each vowel is plotted over time.

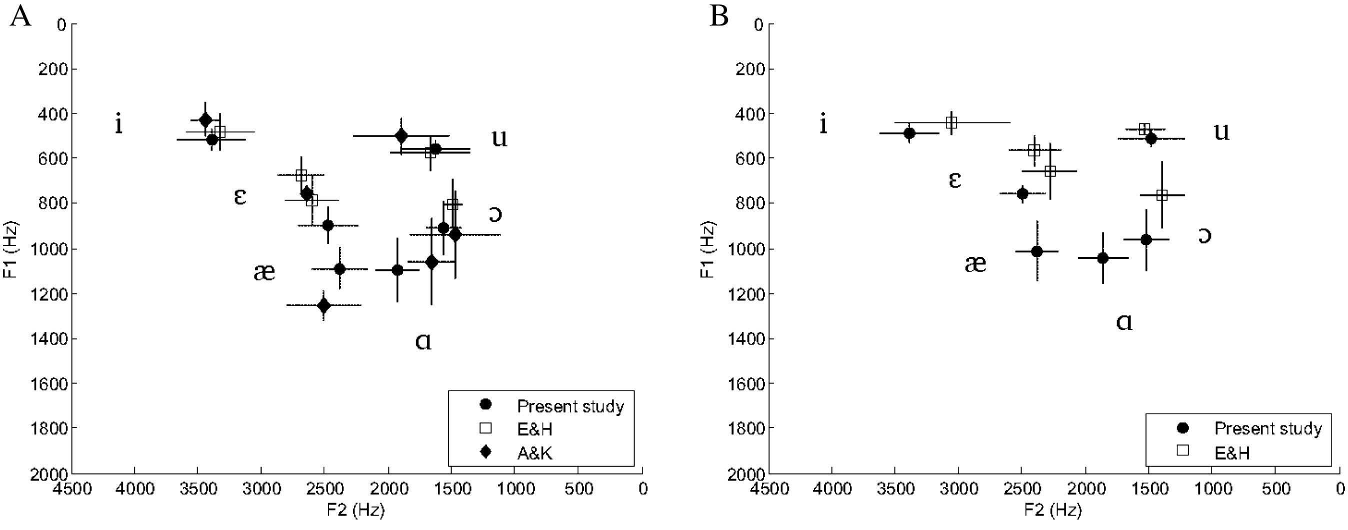
Comparison of the present results to previous studies in which vowels were produced in constant phonetic contexts may highlight important differences between the types of data. Normalized within-subject variability was also calculated by Eguchi and Hirsh (1969) for subjects beginning at age 3 years as well as by Lee et al. (1999) for subjects beginning at age 5 years. The magnitude of normalized within-subject variability is higher in all vowels in the present study than in Eguchi and Hirsh's data. Normalized within-subject variability in the present data of 4-year-old children was higher than in the data of 5-year-old children in Lee et al.'s study.

Dialect and Phonetic Transcriptions

Examination of the vowel space reveals qualitative features among the subjects that could be indicative of dialect region, in particular for the older ages. Figure 4 shows the vowel space of each subject at age 42 months. Dispersion ellipses centered at the mean of each vowel are plotted in the figure; the length of the axes of these ellipses is 1.5 *SD*s in either direction from the mean. Ellipses are labeled at their centers.

Some relative placements of the vowel ellipses could be related to dialect. If we apply the criterion that the ellipse for the vowel /æ/ intersects that for /ɪ/ in order to judge whether the /æ/ is raised and fronted, then there is a high degree of raising and fronting of /æ/ for Rochester, MN (see Figure 4, Panel A). Furthermore, there is a slight degree of raising and fronting of /æ/ for Chicago, IL 1 and Chicago, IL 2

Figure 2. Comparison of formant frequency measurements from the present study at (Panel A) age 36 months, with Eguchi and Hirsh (1969; E&H) and Assmann and Katz (2000; A&K), and at (Panel B) age 48 months, with Eguchi and Hirsh (1969). Mean formant frequencies for each vowel are plotted with 1-*SD* bars in the F1 and F2 directions. The data summarize five to six subjects in the present study, five subjects in Eguchi and Hirsh's study, and three in Assmann and Katz's study. Because Assmann and Katz reported standard errors, *SD*s were calculated from those values by $SD = SE \times \sqrt{3}$, where 3 is the sample size from which the mean is estimated. Error bars indicate the *M* and *SD* between subjects in each study.



(see Figure 4, Panels B and C) compared to the three non-Northern speakers (see Figure 4, Panels D–F). Rochester, MN, and Chicago, IL 2 (see Figure 4, Panels A and C) also have more fronted /ɑ/ than the other speakers. The subject in Panel D of Figure 4 (Oklahoma City, OK) shows an /ɑ/–/ɔ/ merger. In general, /ɪ/ and /ɛ/ overlapped substantially in all subjects except in Figure 4, Panel D (Oklahoma City, OK). The vowel /u/ is more fronted for Oklahoma City, OK, and Logan, UT 2 (see Figure 4, Panels D and F, respectively) than for Rochester, MN, Chicago, IL 1, and Chicago, IL 2 (see Figure 4, Panels A–C).

Comparisons between the phonetic transcriptions of the vowels made by the investigators and the dictionary pronunciation are shown in Table 2. Given the overlap between /ɪ/ and /ɛ/ in the ellipses in Figure 4, it is not surprising that many dictionary /ɪ/s were transcribed as /ɛ/s. Also, there was also a very substantial percentage, 30%, of /ɔ/s transcribed as /ɑ/s. At a lesser level of substitutions or confusions were /ɪ/s for /i/s, at 8%; /ʊ/s for /u/s, at 6%; and /ɑ/s for /ʌ/s, at 4%.

Shapes and Positions of Vowel Spaces

We were interested in comparing the development of each child's vowel space relative to a feature that depends on all the data defining the individual's vowel space. Chung et al. (2012), in a study of five languages, found that language-specific corner vowels are well established in relation to the centroid by age 5 years. The *centroid* is the unweighted mean of vowel locations in log-transformed F1–F2 space. We now consider the shapes of individual vowel spaces as in Chung et al.

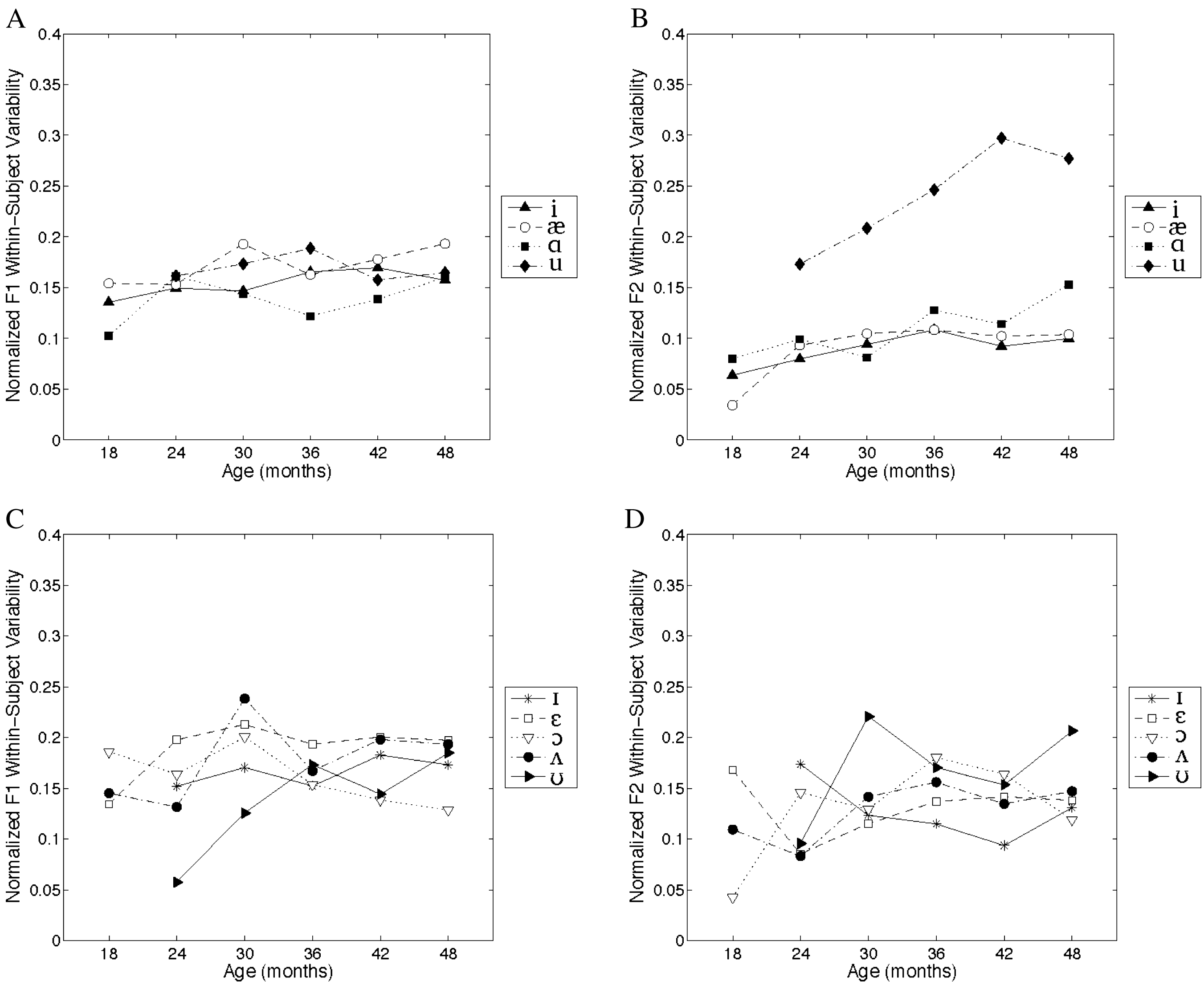
Because the children in the present study did not always produce tokens of every vowel, the location of the centroid was calculated only when most vowels were present, at age 30 months and older. Figure 5 shows the corner vowels and centroids with each vowel space at age 30 months and older superimposed. Age is indicated by symbol shape. The centroid decreases slightly in both F1 and F2 over time for most subjects. Variability in /u/ stands out in several subjects as being larger than in the other corner vowels. The /ɑ/ vowel of Logan, UT 1 (see Figure 5, Panel E) has a substantial trajectory through time.

To better understand the stability or variation in the relation of the vowels to the centroid, we calculated variability from age 30 to 48 months in the direction from the centroid to the vowel. The standard deviation of direction was calculated and multiplied by the mean distance because variability is magnified by distance to the centroid. The results are shown in Figure 6. The back corner vowels, /u/ and /ɑ/, were more variable in direction for some subjects than were the front corner vowels, /i/ and /æ/. The back vowel /ʊ/ was highly variable for some subjects.

Discussion

In this section, we highlight the findings for vowel production when the tokens extracted are largely from recordings taken in conversational settings. We make comparisons with previous studies of vowel production that have used different kinds of elicitation methods. These methods, which we reviewed in the beginning of this article, did not include natural conversational situations. We begin with a discussion of

Figure 3. Normalized within-subject variability for each vowel over time. Panels A and C: F1, Panels B and D: F2. The top two panels show the corner vowels, and the bottom two panels show the non-corner vowels for ease of reading. Normalized within-subject variability is the square root of the mean of the individual variances, divided by the grand mean formant frequency.



within-subject variability because this highlights differences between the current data and the previous data.

Within-Subject Formant Frequency Variability

Comparison of normalized within-subject variability to previous works highlights important differences in the type of data analyzed. Overall levels of normalized within-subject variability in F1 and F2 were much higher in the present study than in the study by Eguchi and Hirsh (1969; ages 3 and 4 years) or by Lee et al. (1999; age 5 years). Furthermore, whereas previous studies have found that normalized within-subject variability decreases with age (Eguchi & Hirsh, 1969; Lee et al., 1999), no apparent decrease in F1 variability, or in F2 variability, was detected over time in the present study (see Figure 3). On the other hand, for each vowel except /u/, overall levels of normalized within-subject variability were lower in F2 than in F1 in the present study as well as in those by Eguchi and Hirsh and Lee et al.

The greater amount of pronunciation variability found in this study compared to earlier ones can likely be attributed to several factors, including variability in segmental phonetic contexts, which would provide greater opportunities for coarticulation, suprasegmental factors arising from speech being spoken in natural contexts, emotional state, and pragmatic factors associated with these meaningful productions,

Figure 4. Vowel spaces of subjects at 42 months. Dispersion ellipses are centered at the mean of each vowel, and axes are 1.5 *SD*s in either direction about the mean in length. Panel A: Northern/Rochester, MN. Panel B: Northern/Chicago, IL 1. Panel C: Northern/Chicago, IL 2. Panel D: Midland/Oklahoma City, OK. Panel E: Western/Logan, UT 1. Panel F: Western/Logan, UT 2. Dialect classifications are from Labov et al. (2006).

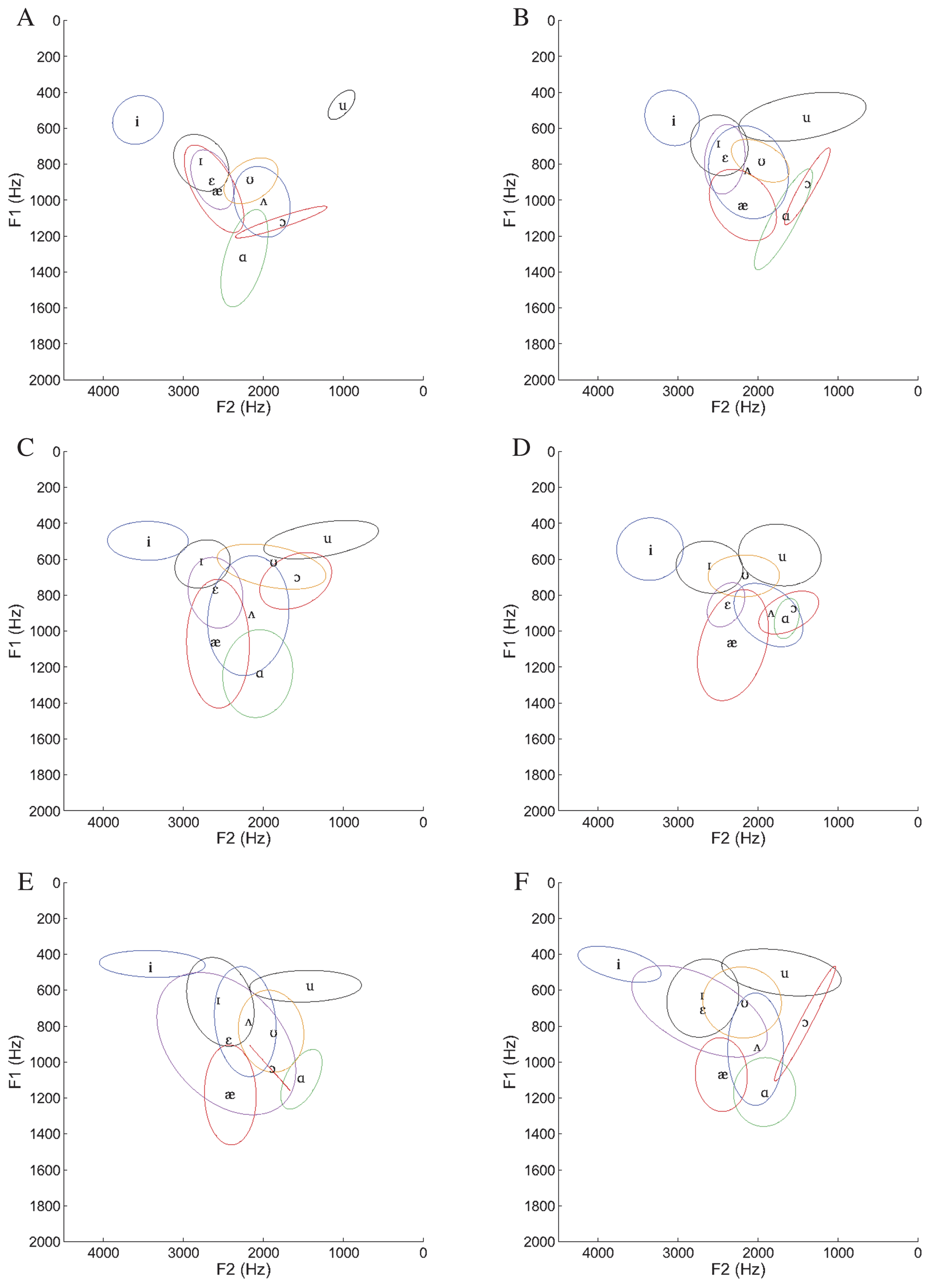
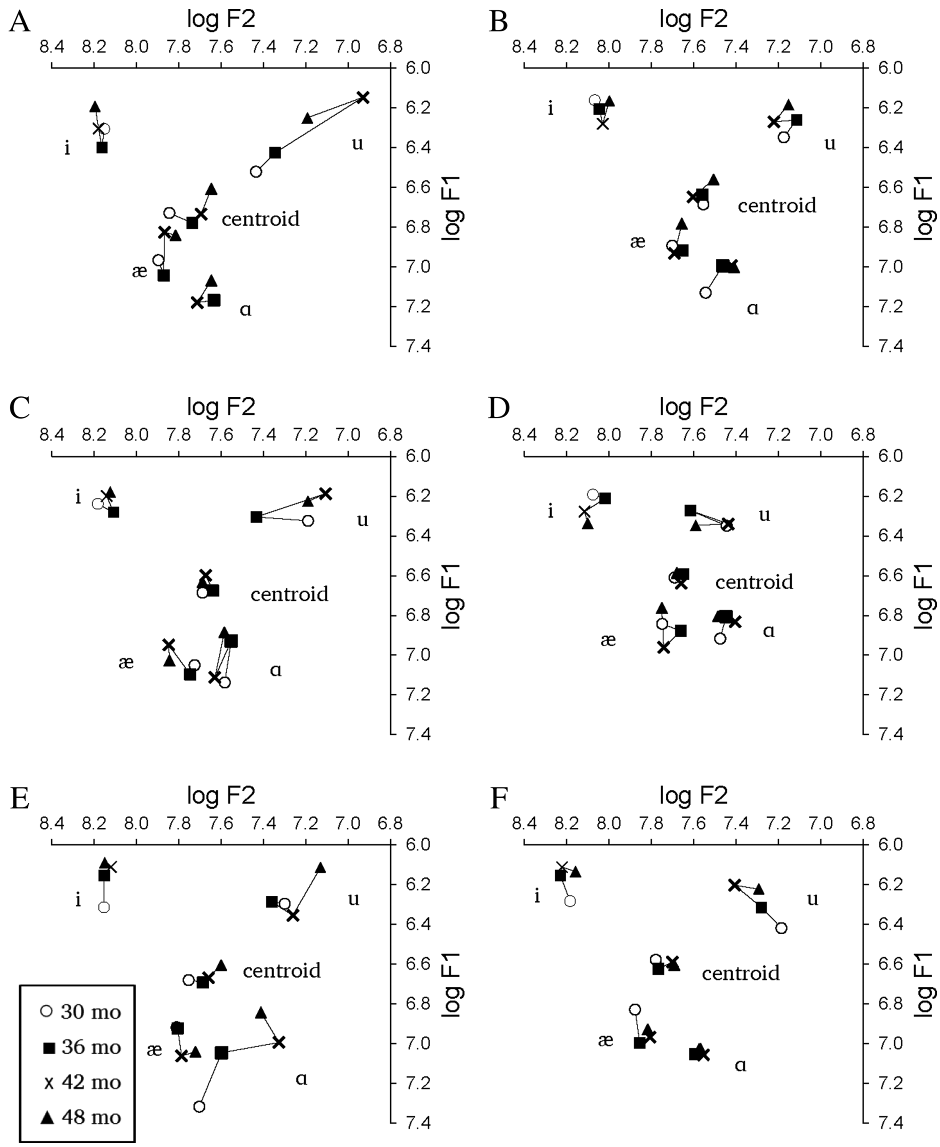


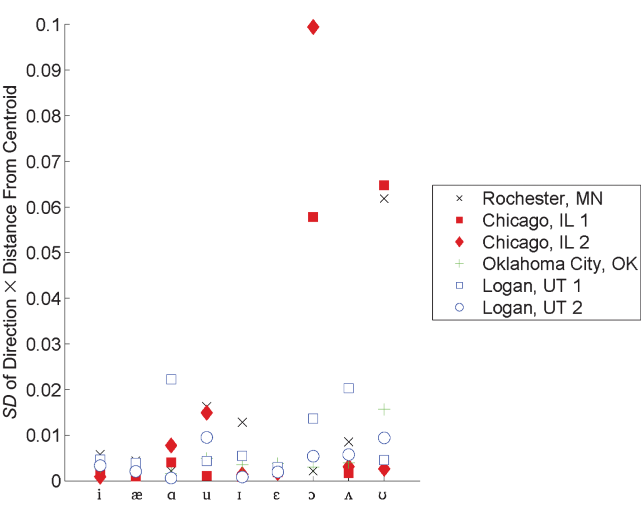
Figure 5. Centroid and corner vowels in log-transformed F1–F2 space, ages 30–48 months. Age is indicated by symbol shape. Panel A: Rochester, MN. Panel B: Chicago, IL 1. Panel C: Chicago, IL 2. Panel D: Oklahoma City, OK. Panel E: Logan, UT 1. Panel F: Logan, UT 2.



as opposed to the elicited forms used by Eguchi and Hirsh (1969) and Lee et al. (1999). Although articulatory precision increases with age, as noted by Eguchi and Hirsh, it appears to be overcome in the present work by the factors just noted as the children's productive vocabulary grows.

The idea that some of the variability can be explained by coarticulation of adjacent phonetic segments is supported by findings from studies with adults (Hillenbrand, Clark, & Nearey, 2001; Stevens & House, 1963). To determine the extent to which coarticulation causes an increase in within-subject variability, a baseline variability—normalized within-subject variability for all vowels following an alveolar consonant—was calculated. The results are shown in Table 4. The percentage of normalized within-subject variability for utterances in alveolar context compared to the within-subject variability in all contexts is less than 85% in F1 for vowels /ʌ/ and /ɑ/ and in F2 for vowels /i/, /ɪ/, /ɛ/, /æ/, /ʌ/, /ɑ/, and /ɔ/. Only the back vowels /ʊ/ and /u/ do not appear to exhibit any additional variability in contexts other than initial alveolar consonant context. In general, Stevens and House's (1963) results indicate that alveolar context perturbs high vowels' F1 more than low vowels' F1 in relation to the other places of articulation, and that is consistent with the results shown in Table 4. Furthermore, it is known that change in F2 from alveolar context is largest for the vowels /ʊ/ and /u/ for adults (Hillenbrand et al., 2001; Stevens & House, 1963). If children behave similarly to adults in this respect, then the variation in F2 in other consonant contexts is smaller

Figure 6. Variability in direction of vowels from the centroid, normalized by distance.



than in the initial alveolar context for these vowels. We conclude that coarticulatory effects, interacting with other factors, make a substantial contribution to the within-subject variability in the present data.

Other factors that interact with phonetic context and contribute to formant variability are suprasegmental and pragmatic factors. For example, Lindblom's (1990) hyper- and hypospeech (H&H) theory, which involves the notion of sufficient discriminability and changes to speech behavior due to pragmatic factors, should have a direct bearing on the variability seen in these children, who were either interacting with a caregiver or an examiner or talking to themselves. Furthermore, in a conversational context, these young children exhibited a variety of emotional states that would certainly influence the acoustics of their vowels.

Formant Frequencies

Formant frequencies measured in the present study are qualitatively similar to those measured at ages 36 and 48 months by Eguchi and Hirsh (1969) and at 36 months by Assmann and Katz (2000), as shown in Figure 2. In particular, F2 measurements are quite similar. We found qualitative differences between the present F1 measurements and those reported by Eguchi and Hirsh in the mid- and low vowels, and in particular a large difference in the F1 of /æ/.

Previous studies (see, e.g., Hillenbrand et al., 1995, who compared formant frequencies from ages 10–12 years to those of Peterson & Barney, 1952) have found differences among data sets that have been attributed to differences in methodology, differences in dialect, or the passage of time. In children's speech, F1 is difficult to measure accurately because of high F0s and the resulting wide spacing of harmonics. The validation of the formant measurements in the present study shows that differences in measurements made independently are about twice as large for F1 than for F2 on a percentage basis. This may be one reason for the differences in F1 reported here and by Eguchi and Hirsh.

Speakers' dialects also may be a factor in the values of the measured formants. The largest difference between the present data and those of Eguchi and Hirsh (1969) is the similarity of /æ/ to /ɛ/ in the latter study. The speakers in that study were from the St. Louis area, which Labov et al. (2006) found is beginning to show the Northern Cities Chain Shift, which we describe in the Dialect section. The majority of speakers in Labov et al.'s study were in their 20s to 40s when the study was conducted in the 1990s, putting them in the same age group as the children in Eguchi and Hirsh's study. The similarity between /æ/ and /ɛ/ was not seen in the speakers from Texas in Assmann and Katz's (2000) study, or in the across-speaker average from different dialect areas in the present study.

Formant Frequencies Over Time

Hierarchical multiple regressions on F1 and F2 were calculated for each subject. Vowel accounted for a large amount of variability in F1 and F2 for all subjects; this simply confirms that vowels differ from one another in F1–F2 space.

The addition of age to the model accounts for substantial additional variability in F1 for three subjects and for additional variability in F2 for two subjects. Each of these subjects showed decreases in the formants, which are expected because of the lengthening of the vocal tract with age. The decreases in F1 are smaller, overall, than the F2 decreases. However, for the subjects who exhibited notable decreases, the percentage decreases given by the regression slopes were 8%, 10%, and 19% for F1, and they were 6% and 7% for F2.

Thus, we found some decrease in formants with time, but not for every subject. In particular, given that the F2 decreases are larger in absolute terms, it is surprising that we did not see an effect in more subjects. One factor that could influence the results of the regressions, which are over all

Table 4. Normalized within-subject variability in F1 and F2 for vowels in alveolar context and the percentage of normalized within-subject variability in alveolar context compared to all contexts.

Variability	i	ɪ	ɛ	æ	ʌ	ɑ	ɔ	ʊ	u
F1	0.142	0.175	0.182	0.150	0.137	0.097	0.120	0.161	0.179
F1%	92	103	91	88	74	69	93	106	106
F2	0.078	0.127	0.128	0.075	0.147	0.101	0.092	0.162	0.236
F2%	50	75	64	44	83	73	72	102	140

vowel tokens, is that the distribution of vowel tokens may shift over time (see the Appendix). Examination of the Appendix indicates that for the Rochester, MN, Chicago, IL 1, and Logan, UT 2 subjects, there were no large changes in the proportions of various vowels. The Chicago, IL 1 and Logan, UT 2 subjects showed decreases in F2. The other subjects, however, exhibited an increase in the number of front vowels as compared to back vowels over time. In other words, the mean F2 may have increased somewhat over time for these subjects. This would be one reason that only two subjects, both with little change in the proportion of front to back vowels, showed decreases in F2 over time.

Furthermore, in this study, we examined vowels on a relatively short time scale; consideration of previous data on a similar time scale is revealing. For example, although Assmann and Katz (2000) found an overall statistically significant decrease of formant frequency over the course of age 3 years to adult, there were not always decreases in formant frequencies over the course of the 2 years between ages 3 and 5, despite selection of vowels for good pronunciation. F1 values increased for /i, ɛ, ɑ/, and F2 values increased or remained stable for /i, ɪ, ɛ, æ, ʌ, ɔ, ʊ/. Lee et al. (1999) also found some increases in formant frequencies for children between age 5 and 6 years, despite an overall trend of decreases between age 5 years and adults. On a short time scale, overall decreases in formant frequencies may not be detected without a significantly larger number of measurements.

Interactions of the individual vowels with age were not included in the present analysis because they resulted in high levels of collinearity; regressions for individual vowels were not computed because more tokens would have been necessary. In order to understand the relationship of formant frequencies to age, Assmann and Katz (2000) conducted an analysis of variance that showed main effects of both vowel and age on the formant frequencies, and which included an Age × Vowel interaction, but this interaction was not significant. In other words, vowels changed over time, but all in a similar manner. Future studies that analyze larger numbers of tokens should consider whether all vowels are similar in their rates of change.

Dialect

Previous studies of children's vowels either have limited the majority of speakers to one region (see, e.g., Eguchi & Hirsh, 1969; Hillenbrand et al., 1995; Lee et al., 1999) or have not considered regional differences (see, e.g., Peterson & Barney, 1952). Comparison of children's vowel spaces from multiple regions does reveal differences that may be due to dialect. Figure 4 shows the vowel space of each subject at 42 months. Three subjects were from the Northern dialect region (see Figure 4, Panels A–C), one was a Midland speaker who lived adjacent to a region of Southern speakers (see Figure 4, Panel D), and two were from the West (see Figure 4, Panels E–F).

One of the most noted features of Northern dialects is the *Northern Cities Chain Shift,* which involves the raising and fronting of /æ/, the lowering and backing of /ɛ/, the lowering and fronting of /ɔ/, and the fronting of /ɑ/, as well as the backing of /ɪ/, /ʌ/, and /ʊ/ (Labov et al., 2006). We observed in the Dialect and Phonetic Transcriptions subsection (in Results section) that the three Northern speakers—Rochester, MN, Chicago, IL 1, and Chicago, IL 2—showed at least a small degree of raising and fronting of /æ/. In comparing the F2 of /ɛ/ to the F2 of /ɑ/, it appears that /ɑ/ can be considered fronted and/or /ɛ/ backed for the Rochester, MN, and Chicago, IL 2 subjects (see Figure 4, Panels A and C). For these Northern speakers, the position of /ɔ/ does not appear to be different from that of the other speakers, except that speaker Chicago, IL 1 (see Figure 4, Panel B) shows more /ɑ/–/ɔ/ merger than expected for a Northern speaker. Also, his /u/ appears to be more fronted than that of the other two Northern speakers. (This speaker heavily yodded his /u/s after alveolar consonants.) There does not appear to be any extra backing of /ɪ/, /ʌ/, and /ʊ/ for any of the three Northern speakers. All three Northern speakers could be showing what Labov et al. (2006) called the *triggering event* for the Northern Cities Chain Shift: the raising and fronting of /æ/. Two of these speakers seem to show the closely related /ɑ/ fronting. Hillenbrand et al. (1995) also found a raised and fronted /æ/ in their older Michigan speakers, which they attributed to the Northern Cities Chain Shift.

Labov et al. (2006) classified Oklahoma City as part of the Midland region. This particular area shows a partial /ɑ/–/ɔ/ merger and fronting of /o/ (we did not study the vowel /o/). The Midland speaker from Oklahoma City (see Figure 4, Pane D) exhibited the most complete /ɑ/–/ɔ/ merger of all of the subjects. This speaker also exhibited /u/ fronting, which is a characteristic of Midland dialect. Oklahoma City is adjacent to a region of Southern speakers. This speaker showed the least overlap in /ɪ/ and /ɛ/, so either the speaker was uninfluenced by the Southern dialect or the separation is due to the fact that nasal contexts were excluded.

The Western dialect features a merger of the vowels /ɑ/ and /ɔ/ and fronting of /u/ (Labov et al., 2006). Only one Western speaker, Logan, UT 2, exhibited possible /u/ fronting (see Figure 4, Panel F). We did not have enough data to determine whether the Logan, UT 1 subject showed an /ɑ/–/ɔ/ merger (see Figure 4, Panel E), and the Logan, UT 2 subject did not appear to show this merger.

In summary, some of the speakers in this study showed dialect features of the region in which they resided. At age 42 months, one could expect that dialect would be greatly influenced by the speech of the caregivers, whose dialect we did not attempt to characterize. This influence may be one reason that some of the children did not exhibit features of their local dialect.

Shape and Position of the Vowel Space

The location of the centroid in log-transformed F1–F2 space was found and plotted in Figure 5. The centroid formants decreased over time somewhat, but they provide a well-defined reference against which to compare the vowels. The corner vowels showed some variability in their relation to the centroid, but no consistent trends emerged. This suggests that there was no rearrangement of the vowel space in relation to the centroid. The relation of the corner vowels to one another appears to have been established early in the development of the vowel space, even as formant frequencies continued to decrease, as an expansion of vocabulary and phonetic contexts occurred, and as development of suprasegmental and pragmatic competence contributed to a large amount of variability in the realization of the vowels.

We considered the variability of each vowel in the vowel space by quantifying the relation of the vowels to the centroid. Variability in direction from the centroid, shown in Figure 6, reveals some differences in the vowels. The corner vowels were consistently located in relation to the centroid in terms of direction, with the exception of one speaker, who had a large amount of variability in direction for /ɑ/. Within the set of four corner vowels, the front vowels were the most consistent. The non-corner back vowels tended to be more variable in direction to the centroid. Ménard et al. (2007) measured larger dispersion ellipses (in Bark) for back vowels compared to front vowels for French-speaking 4- and 8 year-old children. This difference in area was found to be larger for the children compared to the adults.

Conclusion

In this study, we measured F1 and F2 of children's vowels in a naturalistic context at the youngest ages at which they are acquiring a vocabulary. The results compare well with previous such studies of older children's speech. The vowel spaces look qualitatively similar to those found for older children in previous work, but the present study also highlighted interspeaker dialect differences not previously examined in children under age 4 years.

Within-subject variability in young children's vowel production in these mostly conversational environments is higher than that in previous studies. Furthermore, it remained relatively constant through the study, compared with the previous studies that have examined elicited speech, in which consistency increased with age. That conversational speech in young children produced more within-speaker variability than elicited forms is not surprising. That this variability does not appear to change with age is an important finding. Variability can be caused by coarticulation with surrounding phonetic segment as well as the important factors of intonation and prosody, such as sentence focus. The consequences of pragmatic factors in variability should also be taken into account. It appears that the increase in the number of phonetic segmental contexts, as well as possible increases in the intonational and prosodic environments in which the vowel segments are embedded, counteracts any increase in pronunciation precision with age found in citation form stimuli. An interesting question to consider is whether, as the results of previous studies suggest, formant frequency within-subject variability in adults is also lower than in children when the variety of contexts found in conversational speech is considered.

There were some substantial decreases in overall F1 and F2 for some subjects over the span of 30 to 48 months. According to magnetic resonance imaging data from Vorperian and colleagues (Vorperian, Kent, Gentry, & Yandell, 1999; Vorperian et al., 2005), the vocal tract increases rapidly in length from approximately 7.5 cm (or 47% of adult size) in a newborn to 9.5 cm (or 62% of adult size) at age 18 months. The vocal tract then continues to steadily lengthen somewhat more slowly to attain about 68% of adult size at 48 months of age (Vorperian et al. 2005, Figure 4), which means that there is a 10% increase in average vocal tract length between 18 and 48 months. From 30 to 48 months, the increase is even smaller, at about 5%. Overall, this increase in vocal tract length would mean that, at 48 months, formant frequencies are reduced by 95% of their values at 30 months, if vocal tract length were the only factor. For the subjects for whom changes in F1 or F2 were detected, the reduction was always greater than 5%, sometimes by a substantial amount. With such small changes due to vocal tract lengthening, it is not surprising that we were unable to systematically detect the acoustic effect of length changes that occurred between ages 30 and 48 months for all subjects. These considerations are based on group averages, so some children will exhibit even less than 5% vocal tract lengthening. Furthermore, it is not possible from these data to infer the reasons that some subjects do not show decreases in formant frequencies that would be expected with the growth of the vocal tract. Only the fact that the vowel inventory changed for some of the speakers can be offered as a reason, but this does not discount other possible causes, including social influences.

Despite the many changes occurring in children's speech over the course of 30 to 48 months, the relation of the corner vowels to the centroid in the vowel space remains a qualitatively constant feature. Although further changes, such as lowering of formant frequencies and improvement in pronunciation precision, are expected beyond age 48 months, the overall relations of vowels to one another may be well established at a young age. However, the differences in variability between back and front vowels for these children warrants further research.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Grants R01 DC001247 (awarded to CReSS LLC) and R01 DC006237 (awarded to the fourth author).

References

***American heritage college dictionary* (3rd ed.).** (1993). Boston, MA: Houghton-Mifflin.

Assmann, P. F., & Katz, W. F. (2000). Time-varying spectral change in the vowels of children and adults. *The Journal of the Acoustical Society of America, 108,* 1856–1866.

Bickel, P. J., & Doksum, K. A. (1977). *Mathematical statistics: Basic ideas and selected topics*. San Francisco, CA: Holden-Day.

Chung, H., Kong, E. J., Edwards, J., Weismer, G., Fourakis, M., & Hwang, Y. (2012). Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America, 131,* 442–454.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Oto-Laryngologica Supplementum, 257,* 307–356.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, United Kingdom: Sage.

Gilbert, H. R., Robb, M. P., & Chen, Y. (1997). Formant frequency development: 15 to 36 months. *Journal of Voice, 11,* 260–266.

Hays, W. L. (1981). *Statistics* (3rd ed.). New York, NY: Holt, Rinehart and Winston.

Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America, 109,* 748–763.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97,* 3099–3111.

Kent, R. D., & Forner, L. (1979). Developmental study of vowel formant frequencies in an imitation task. *The Journal of the Acoustical Society of America, 65,* 208–217.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America, 72,* 353–365.

Kent, R. D., Osberger, M. J., Netsell, R., & Hustedde, C. G. (1987). Phonetic development in identical twins differing in auditory function. *Journal of Speech and Hearing Research, 52,* 64–52.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change*. New York, NY: Mouton de Gruyter.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America, 105,* 1455–1468.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, the Netherlands: Kluwer Academic.

McGowan, R. S., Nittrouer, S., & Chenausky, K. (2008). Speech production in 12-month-old children with and without hearing loss. *Journal of Speech, Language, and Hearing Research, 51,* 879–888.

Ménard, L., Schwartz, J.-L., Boë, L.-J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics, 35,* 1–19.

Nittrouer, S. (2010). *Early development of children with hearing loss*. San Diego, CA: Plural.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120–132.

Nittrouer, S., Studdert-Kennedy, M., & Neely, S. T. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research, 39,* 379–389.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America, 109,* 2988–2998.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Smith, A. (2010). Neural control of orofacial movements in speech. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *Handbook of phonetic sciences* (2nd ed., pp. 251–296). Chichester, United Kingdom: Wiley-Blackwell.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research, 6,* 111–127.

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research, 50,* 1510–1545.

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of the vocal tract structures: Preliminary results. *International Journal of Pediatric Otorhinolaryngology, 49,* 197–206.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America, 117,* 338–350.

Wilcox, K., & Morris, S. (1999). *Children's Speech Intelligibility Test*. San Antonio, TX: The Psychological Corporation.

Appendix

Number of Vowels Measured for Each Subject, at Each Age

Age (mos)	Subject	i	ɪ	ɛ	æ	ʌ	ɑ	ɔ	ʊ	u
18	Rochester, MN	4	0	0	0	0	0	0	0	0
	Chicago, IL 1	2	1	0	1	0	0	0	0	0
	Chicago, IL 2	2	0	0	0	7	0	7	0	1
	Oklahoma City, OK	1	0	0	0	2	0	1	0	0
	Logan, UT 1	4	0	4	1	1	0	0	0	0
	Logan, UT 2	4	0	1	2	0	32	0	0	0
	Total	17	1	5	4	10	32	8	0	1
24	Rochester, MN	3	2	4	2	2	0	0	0	2
	Chicago, IL 1	37	12	28	6	26	3	2	2	9
	Chicago, IL 2	9	2	14	6	7	1	0	2	5
	Oklahoma City, OK	20	3	2	14	0	3	0	0	0
	Logan, UT 1	8	2	4	2	0	2	6	0	1
	Logan, UT 2	11	19	1	6	1	0	0	1	0
	Total	88	40	53	36	36	9	8	5	17
30	Rochester, MN	18	6	4	9	3	0	0	2	8
	Chicago, IL 1	68	39	37	25	18	6	9	3	22
	Chicago, IL 2	13	11	5	28	6	4	0	5	15
	Oklahoma City, OK	28	14	11	15	4	6	6	2	6
	Logan, UT 1	11	15	3	9	7	0	0	0	7
	Logan, UT 2	30	49	6	9	6	1	0	0	5
	Total	168	134	66	95	44	17	15	12	63
36	Rochester, MN	29	29	9	14	13	3	1	5	12
	Chicago, IL 1	56	76	18	19	28	13	10	13	20
	Chicago, IL 2	34	55	9	22	21	7	3	8	13
	Oklahoma City, OK	16	21	12	27	4	8	5	3	11
	Logan, UT 1	11	24	9	11	8	2	2	0	10
	Logan, UT 2	37	50	21	30	10	5	0	13	23
	Total	183	255	78	123	84	38	21	42	89
42	Rochester, MN	16	29	6	11	6	4	3	4	3
	Chicago, IL 1	57	85	20	25	26	18	5	8	26
	Chicago, IL 2	52	63	11	42	12	24	5	18	32
	Oklahoma City, OK	38	73	6	37	14	9	6	23	10
	Logan, UT 1	27	45	10	13	17	4	2	4	7
	Logan, UT 2	42	52	18	53	24	9	3	13	37
	Total	232	347	71	181	99	68	24	70	115
48	Rochester, MN	36	33	7	19	12	4	4	8	13
	Chicago, IL 1	61	72	23	20	24	8	10	12	20
	Chicago, IL 2	60	61	21	25	20	16	6	10	29
	Oklahoma City, OK	61	50	30	35	28	23	20	9	13
	Logan, UT 1	45	47	11	30	14	6	1	3	20
	Logan, UT 2	56	100	10	48	35	10	11	18	38
	Total	319	363	102	177	133	67	52	60	133