Acoustic development of vowel production in native Mandarin-speaking children

Jing Yang
Communication Sciences and Disorders,
University of Central Arkansas
jyang@uca.edu

Robert A. Fox
Department of Speech and Hearing Science,
The Ohio State University
fox.2@osu.edu

The present study aims to document the developmental profile of static and dynamic acoustic features of vowel productions in monolingual Mandarin-speaking children aged between three and six years in comparison to adults. Twenty-nine monolingual Mandarin children and 12 native Mandarin adults were recorded producing ten Mandarin disyllabic words containing five monophthongal vowel phonemes /a i u y ɤ/. F1 and F2 values were measured at five equidistant temporal locations (the 20–35–50–65–80% points of the vowel's duration) and normalized. Scatter plots showed clear separations between vowel categories although the size of individual vowel categories exhibited a decreasing trend as the age increased. This indicates that speakers as young as three years old could separate these five Mandarin vowels in the acoustic space but they were still refining the acoustic properties of their vowel production as they matured. Although the tested vowels were monophthongs, they were still characterized by distinctive formant movement patterns. Mandarin children generally demonstrated formant movement patterns comparable to those of adult speakers. However, children still showed positional variation and differed from adults in the magnitudes of spectral change for certain vowels. This indicates that vowel development is a long-term process which extends beyond three years of age.

1 Introduction

In contrast to the large number of studies on consonants, relatively few studies focus on children's vowel acquisition and development, especially in non-English-speaking children. This is probably due to the claims in previous research that vowels are easier to acquire than consonants and are acquired at an early stage of childhood (Waterson 1971, N. V. Smith 1973, Zhu & Dodd 2000). However, vowel acquisition is not as straightforward as it has been thought. Although a large number of studies on English-acquiring infants consensually suggest that the corner vowels /ɑ i u/ are learned earlier than other vowels and vowel acquisition is generally complete prior to three years of age, they report varied timelines

doi:10.1017/S0025100317000196

of acquisition of individual vowels (e.g. Templin 1957, Hare 1983, Paschall 1983, Davis & MacNeilage 1990, Stoel-Gammon & Herrinton 1990). Moreover, even typically developing children exhibit great individual differences in phonological development (Vihman, Ferguson & Elbert 1986, Vihman 1993). Thus, the process of vowel development deserves further examination. Research on a wider variety of languages is also needed.

A majority of existing studies on vowel development have focused on English-speaking children. An early study of Wellman et al. (1931) found that vowels /ɪ æ ʊ/ were produced with the lowest accuracy rate while vowels /i u/ were produced with the highest accuracy rate by children across all age groups (ages from two years to six years). Templin (1957) reported that over 90% of vowels and diphthongs in words were produced correctly in children at three years of age. Since then, many studies examined vowel acquisition in younger children. For example, Paschall (1983) recorded spontaneous speech from 20 infants aged between 16 and 18 months and found that children at this age produced vowels /ɑ u i/ with the highest accuracy and vowels /e ɛ ɝ aɪ/ with the lowest accuracy. Davis & MacNeilage (1990) traced vowel acquisition from one child over the period from 14 to 20 months. They found that the vowels /u o i ɪ/ were produced with the highest accuracy while the vowels /ɚ ɛ/ were produced with very low accuracy. Otomo & Stoel-Gammon (1992) examined the acquisition of unrounded vowels /i ɪ ɛ æ ɑ/ in six normally developing English children between 22 and 30 months of age. They found that among the five vowels, /i ɑ/ were acquired earlier and with higher accuracy than /ɪ ɛ/.

Compared to the well documented development of vowel acquisition in English-speaking children, there is relatively little research on children acquiring Mandarin. The limited findings of vowel acquisition in Mandarin children are primarily based on longitudinal case studies using the traditional approach of transcription-based accuracy ratings (Jeng 1979, Zhu & Dodd 2000, Si 2006,). Subsequently, Shi & Wen (2007) examined the acquisition of seven Mandarin vowels /a i u y ə ɹ̪̩ ɻ̩/[1] by 40 native Mandarin-speaking children between one and six years old. On the basis of evaluating the accuracy rate of the children's vowel production, Shi & Wen summarized the acquisition order of Mandarin vowels a > i > ə > u > ɻ̩ > ɹ̪̩ > y. In addition, Shi & Wen (2007) provided several samples of individual children's acoustic vowel space to characterize the nature of their speech errors. However, except for these sporadic efforts directed at establishing an acoustic profile for individual Mandarin-speaking children's vowel production, no study has systematically examined the acoustic properties of vowel production in native Mandarin-speaking children at different ages.

The purpose of the present study is to expand our knowledge of children's vowel development by systematically examining the phonetic-acoustic characteristics of vowel production in native Mandarin-speaking children. In particular, the research aims are twofold:

(i) to document the developmental profile of static and dynamic acoustic features of five basic vowels in Mandarin-speaking children between three and six years of age, and
(ii) to examine the extent to which the acoustic characteristics of children's vowels are similar to or different from those of Mandarin-speaking adults.

A substantial body of previous research has investigated the developmental change of acoustic features in native English-speaking children across different age spans (e.g. Kent & Murray 1982, Busby & Plant 1995, Kuhl & Meltzoff 1996, Gilbert, Robb & Chen 1997, Lee, Potamianos & Narayanan 1999). Children's vocal tracts grow as a function of their chronological age. It is well known that formant frequencies are inversely related to the length of vocal tract. Correspondingly, children's formant frequencies and F1-F2 vowel space area decrease as a function of chronological age (Vorperian & Kent 2007). Collectively,

---

[1] Shi & Wen (2007) used /ə/ to represent the close-mid back unrounded vowel /ɤ/. They used Chinese phonetic symbols [ɿ] and [ʅ] for the Mandarin high front apical vowel and high back apical vowel. The present study substituted [ɿ] and [ʅ] with the IPA equivalents [ɹ̪̩] and [ɻ̩] proposed by Lee-Kim (2014).

these studies on the acoustic properties of children's vowel production have provided a fairly complete profile of acoustic development of vowel production from infancy to early adulthood, at least for English-speaking children. However, it remains unclear if the effects produced by physical changes on the vocal tract are removed, whether the acoustic features of the children's vowels (e.g. formant patterns and formant dynamics) will be the same as those of adults. If not, it suggests that in addition to the lengthening of vocal tract, there may be other factors accounting for the difference of vowel acoustic features between children and adults, indicating that the child at this particular age has not fully acquired the vowels yet.

To date, most studies examining the acoustic properties of children's vowel productions focused on the midpoint value of a vowel's formants and its position in the F1 × F2 acoustic vowel space. The midpoint values have been assumed to represent the steady state portion of the vowel. However, a growing body of literature shows ample evidence that vowels are more precisely characterized by inherent spectral change (Neary & Assmann 1986, Harrington & Cassidy 1994, Hillenbrand et al. 1995, Assmann, Nearey & Bharadwaj 2013, Neary 2013, Yuan 2013). Vowel spectral change also conveys important information associated with the speaker's dialect (Fox & Jacewicz 2009) and generation (Jacewicz, Fox & Salmons 2011a). Furthermore, vowel spectral change is also related to a child's development of speech motor control. Both temporal and spectral measures have long been used as indices to estimate speech motor development in children (e.g. B. L. Smith 1978, 1991; Kent & Forner 1980; Nittrouer 1993). The apparent importance of vowel spectral change motivates us to further explore the dynamic features of vowel production by children to better understand the nature of their speech sounds. Only a few studies examined the development of vowel dynamic spectral change in English-speaking children (Jacewicz, Fox & Salmons 2011b, Assmann et al. 2013). These studies have provided basic knowledge regarding the acoustic features of vowel spectral change in native English-speaking children older than five years of age. However, little research has been done to explore the pattern of spectral change in children from younger age groups and/or different language backgrounds. In addition, except for a few studies providing a quantitative description of vowel spectral change for native English adults (e.g. Fox & Jacewicz 2009), very little has been done to measure the magnitude of vowel spectral change in children, especially young children, which is vital to define the characteristics of children's formant movement.

According to previous findings on vowel acquisition in English-speaking and Mandarin-speaking children, we predicted that – following a general and perhaps universal trend in vowel acquisition – Mandarin-speaking children would show earlier development of acoustic features for the corner vowels /a i u/ than other vowels. In particular, we expected that the early developed vowels would approximate the adult targets to a greater extent than the later developed vowels. In addition, according to previous studies on the compatible vowel spectral dynamic patterns between relatively old English-speaking children and adults, we predicted that the older group of Mandarin-speaking children would show spectral dynamic patterns similar to the adults. Finally, we explored whether the vowel dynamics of the younger group of children are similar to the adults.

2 Method

2.1 Speakers

The speakers included 29 native Mandarin-speaking children (13 females and 16 males) aged between three and six years, and 12 native Mandarin-speaking adults (six females and six males) aged 23–58 years. These speakers fell into three age groups: younger children (AY: 3–4;11 (age in years;months), 14 children, eight males and six females), older children (AO: 5–6;11, 15 children, eight males and seven females) and adults (AA). All children were

Table 1 Word list of Mandarin vowels used in production data collection.

Vowel	Pinyin	IPA	Gloss
a	dà xiàng	/taɕiɑŋ/	elephant
	dà suàn	/tasuan/	garlic
i	pí qiú	/pʰitɕʰioʊ/	ball
	bí zi	/pitsɹ̺/	nose
u	tù zi	/tʰutsɹ̺/	rabbit
	pú tao	/pʰutʰaʊ/	grape
y	jú zi	/tɕytsɹ̺/	orange
	yú tóu	/ytʰoʊ/	fish head
ɤ	gē ge	/kɤkɤ/	older brother
	gē zi	/kɤtsɹ̺/	pigeon

born in Mandarin-speaking regions and raised in the Beijing area, where the data collection occurred. Both parents of all children spoke Mandarin in daily life. All adults were born and raised in Mandarin-speaking regions and were also recruited from the Beijing area. Five of the twelve adults were teachers employed in the kindergarten where the child speakers were recruited. No speech or language disorders were reported by any of the participants or their caregivers.

2.2 Speech materials

The speech materials (see Table 1) included ten Mandarin disyllabic words (e.g. *dà suàn* /tasuan/ ‘garlic’, *pí qiú* /pʰitɕʰioʊ/ ‘ball’) containing five Mandarin monophthongal vowel phonemes /a i u y ɤ/. Unlike Shi & Wen (2007), the present study did not include the apical allophones of /i/, the high front apical vowel [ɹ̺] and the high back apical vowel [ɻ] (represented as [ɿ] and [ʅ], respectively, in Chinese linguistics). These two vowels occur after alveolar and retroflex affricates/fricatives /ts tsʰ s/ and /tʂ tʂʰ ʂ/, respectively. The high front laminal vowel [i] occurs in all other consonant environments. All words were selected on the basis of their picturability and familiarity to young children. Each vowel occurred in two different words and each word was repeated twice. In each disyllabic word, the target vowel always appeared in the first syllable which started with a stop consonant, with the exception of the word for ‘fish head’ (see Table 1), which started with the vowel /y/; this was due to the phonotactic constraints in Mandarin syllable structure. The target syllable was always CV. The initial consonant in the second syllable was not strictly controlled. In addition, the tone environment of the target syllable was not strictly controlled except that tone 3 was avoided to reduce the effect of tone–vowel interactions on the vowel (e.g. f0, vowel duration) which are primarily manifested with tone 3 (Lin 1988, Xu, Tsai & Pfingst 2002, Hoole & Hu 2004, Chang 2010).

2.3 Procedures

The speech samples were collected through a visual-auditory word repetition task in a quiet room under the control of a custom MATLAB program. During the recording period, each participant was seated in front of a laptop computer wearing a Shure SM10A head-mounted microphone situated approximately one inch from the speaker’s mouth. For each participant, the recording session was divided into two blocks. In each block, pictures corresponding to the 10 target words were randomly ordered and presented on the computer screen. The same random order was used for both recording blocks. For each word, the participants were presented with an audio prompt that was produced in a citation form by a native Mandarin speaker. The participant was then asked to repeat the word once immediately after the prompt. In contrast to spontaneous speech elicitation, this method of data collection ensured that the speech samples were the specific target word as expected and were collected under better controlled conditions (Edwards & Beckman 2008). Speech samples were recorded directly onto a hard disk of the laptop with a 16-bit quantization rate and 44.1 kHz sampling rate.

2.4 Acoustic measurements

2.4.1 Formant frequencies

All tokens were down-sampled to 11.025 kHz using a custom MATLAB program prior to acoustic analysis. The spectrographic analysis program TF32 (Milenkovic 2003) was then used to extract the frequency values of the first two formants, F1 and F2. Following previous studies (Fox & Jacewicz 2009, Jacewicz, Salmons & Fox 2009, Jacewicz, Fox & Salmons 2011c), formant frequencies were measured at five equidistant temporal locations (the 20–35–50–65–80% points) to capture the dynamic spectral change in the vowel. The 20% and 80% points were selected over the initial and final points in order to reduce the influence (in the form of formant transitions) of the preceding and following consonants. The temporal locations were determined with reference to the landmark locations of vowel onset and offset using standard measurement criteria (Kent & Read 1992). In particular, vowel onset was defined as the initial zero crossing of the first period of voicing following stop closure release or cessation of frication. For the word /y/ (*yú* 'fish') which is composed of one single vowel with no initial consonant, the onset of the vowel was defined as the start of first clear period of voicing. Vowel offset was set at the zero crossing point of the final period of voicing.

2.4.2 Vowel normalization

As the purpose of the present study was to investigate age-related differences of vowel acoustic measurements in Mandarin-speaking children after the effect of vocal tract length among speakers was eliminated, normalized formant frequency values were calculated using the method in Lobanov (1971). Lobanov's method is a vowel-extrinsic normalization approach which has been found to be one of the most effective normalization methods (Adank, Smits & van Hout 2004). It converts formant frequency values into z scores through the formula

$$F_{n[V]}{}^{N} = \left(F_{n[V]} - \mathrm{MEAN_n}\right)/\mathrm{S_n} \tag{1}$$

where $\mathrm{F}_{n[V]}{}^{\mathrm{N}}$ is the normalized value of formant n for the vowel V, and $\mathrm{MEAN_n}$ and $\mathrm{S_n}$ are the mean and standard deviation of formant n of the target speaker (Thomas & Kendall 2007). The normalized formant frequency values of each speaker were then rescaled into Hz-like values using the method proposed by Thomas & Kendall (2007) to facilitate interpretation of normalized vowel spaces. In particular, the equations used for rescaling were:

$$\mathrm{F'_1} = 250 + 500\left(\mathrm{F^N}_1 - \mathrm{F^N}_{\mathrm{1MIN}}\right)/\left(\mathrm{F^N}_{\mathrm{1MAX}} - \mathrm{F^N}_{\mathrm{1MIN}}\right) \tag{2}$$

$$\mathrm{F'_2} = 850 + 1400\left(\mathrm{F^N}_2 - \mathrm{F^N}_{\mathrm{2MIN}}\right)/\left(\mathrm{F^N}_{\mathrm{2MAX}} - \mathrm{F^N}_{\mathrm{2MIN}}\right) \tag{3}$$

where $\mathrm{F'}_i$ is a rescaled, normalized formant; $\mathrm{F^N}_\mathrm{i}$ is a Lobanov normalized formant value; and $\mathrm{F^N}_{i\mathrm{MIN}}$ and $\mathrm{F^N}_{i\mathrm{MAX}}$ are the minimum and maximum values of Lobanov normalized $\mathrm{F^N}_i$ across the entire dataset.

2.4.3 Trajectory length

To quantify the amount of the formant movement in a vowel, trajectory length (TL) was calculated on the basis of the rescaled normalized F1 and F2 values at five time locations in the acoustic vowel space (Fox & Jacewicz 2009). TL was defined as the sum of the Euclidean distances (in Hz) between each two consecutive temporal points (i.e. 20–35%, 35–50%, 50–65%, 65–80%),

$$TL = \sum_{n=1}^{4} VSL_n \tag{4}$$

where the length of each vowel section (VSL) was calculated using the formula:

$$VSL_n = \sqrt{(F1_{n+1} - F1_n)^2 + (F2_{n+1} - F2_n)^2} \tag{5}$$

The assumption was that a longer trajectory length indicates a greater change of formant frequency values and greater magnitude of formant movement. This measure provided a detailed assessment of the magnitude of formant movement over the course of vowel duration particularly for the curved formant tracks.

2.5 Statistical analysis

Each acoustic measure was subject to ANOVA tests. For vowel duration, midpoint formant frequency values, and vowel trajectory length, a two-way repeated-measures ANOVA was conducted with vowel as the within-subject factor and age group as the between-subject factor. When a main effect from the ANOVAs was significant, a post hoc test (Bonferroni with adjustment for multiple comparisons) was selected for pair-wise comparisons. For the vowel space area, one-way ANOVA was conducted to compare the difference among the three age groups. For vowel trajectory length, the analysis of probability density function was adopted in addition to ANOVA tests.

3 Results

3.1 Vowel duration

Vowel duration was determined from the manually located onset and offset landmarks. As shown in Figure 1, vowels produced by children were generally longer than those produced by adults, consistent with the findings in previous studies (e.g. B. L. Smith 1978, Barton & Macken 1980, Lee et al. 1999). However, there was no obviously greater variation in either group of children as compared to the adults. The results of a two-way repeated-measures ANOVA revealed a significant main effect of vowel ($F(4,152) = 16.3, p < .0001, \eta_p^2 = .300$) and age ($F(2,38) = 4.313$, $p = .021$, $\eta_p^2 = .185$). No vowel by age interaction was found, indicating that children followed the vowel duration pattern of adults. The Bonferroni post hoc test showed that both younger (AY) and older children (AO) produced significantly longer vowel durations than the adults (AA). However, no significant difference was found between the two groups of children.

3.2 Midpoint F1 by F2 vowel space

Traditionally, midpoint formant frequency values have been used to characterize the basic acoustic feature of vowels (and their position in the F1 × F2 plane). In the present study, the rescaled normalized F1 and F2 at midpoint location were used to plot the vowel dispersion pattern (shown in Figure 2). Vowel ellipses were plotted in a manner similar to that used in Zhou & Xu (2008) for tonal ellipses. In particular, for each vowel ellipse, the rescaled normalized F1 × F2 scatter plot was fitted linearly and the positive angle of the linear fit was taken as the direction of the semimajor of the ellipse. The mean of the rescaled normalized F2 values along the semimajor axis was defined as the ellipse center. A line perpendicular to

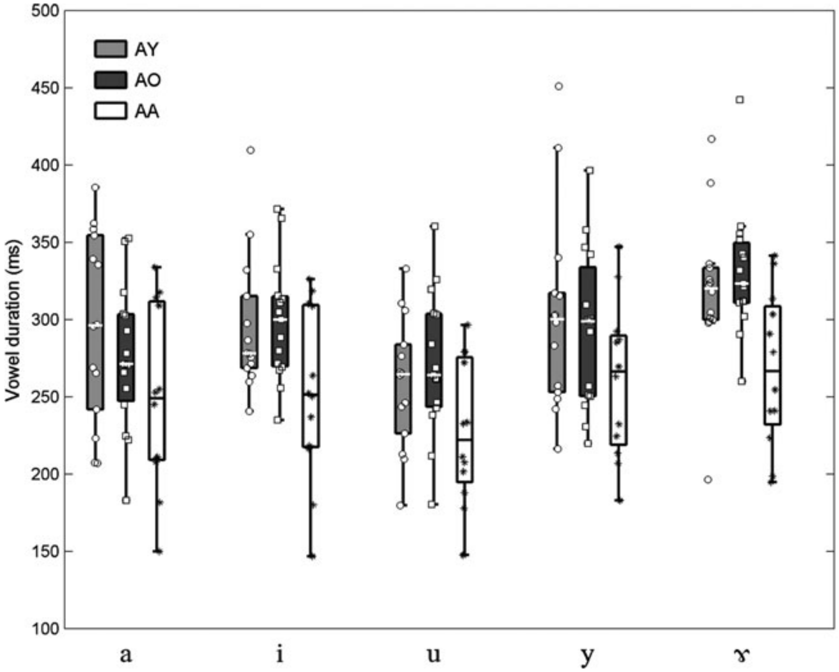

Figure 1 Box plot showing the vowel duration for each Mandarin vowel in each age group. The dots in the boxes represent the vowel duration of individual speakers within each group. Each box shows horizontal lines at the lower quantile, median and upper quartile values. The whiskers show the range of the data.

the semimajor axis at the center was defined as the semiminor axis. The lengths of the two axes were defined as two standard deviations of all data points away from the center along the respective axes. Each ellipse thus incorporated approximately 95% of the data points for each vowel category.

As shown in the plot for adults (right panel), the speech samples were highly concentrated in five clearly separated vowel phoneme categories. The vowels /a i u/ were located in the corner positions. /y/ was close to /i/ while /ɤ/ was located in a relatively central position. These vowel categories were also clearly separated in both groups of children. However, although the children differentiated the five vowel categories acoustically, the sizes of the ellipses decreased as a function of chronological age. The area of each ellipse based on the rescaled normalized formant frequency values for each vowel category in each age group was calculated and is shown in Table 2 – the areas of the ellipses decreased as a function of speakers' age for all five vowel categories. Thus, although the children could separate these five vowel categories in the acoustic space, they demonstrated greater variability than the adults and the variability decreased as the children grew older.

Figure 3 shows the means and standard deviations of the rescaled normalized formant frequency values at midpoint location for each Mandarin vowel in each group of speakers. It can be seen that both groups of children deviated from adult norms for the five Mandarin vowels – the deviation residing in the F2 dimension for most vowels. For the vowel /i/, both the younger and older children produced higher F2 than did the adults. For the vowel /y/, both younger and old children produced lower F2 than did the adults. For the vowel /a/, the older children produced higher F2 than did the adults. For the vowel /ɤ/, the older children produced lower F2 than did the adults. A two-way repeated-measures ANOVA was used to

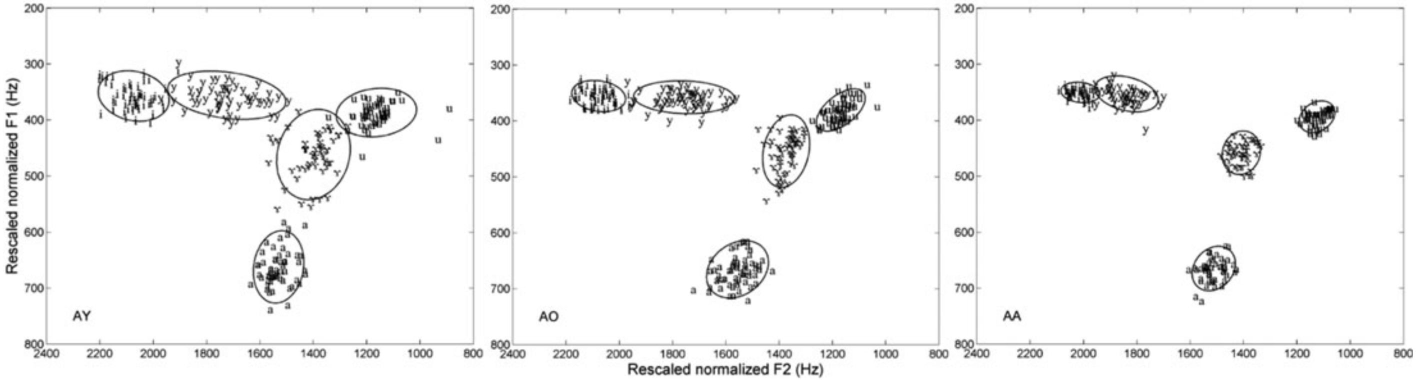

Figure 2 Scatter plots of midpoint formant frequency values for individual Mandarin vowel categories in the younger children (left), the older children (middle) and the adults (right). Each data point represents one vowel token.

Table 2 The area of ellipse of each Mandarin vowel in each age group (in Hz^2).

Age group	a	i	u	y	ɤ
AY	19,097	18,581	20,376	28,841	34,642
AO	18,062	8,777	8,968	17,408	17,591
AA	9,861	4,122	5,699	11,373	8,733

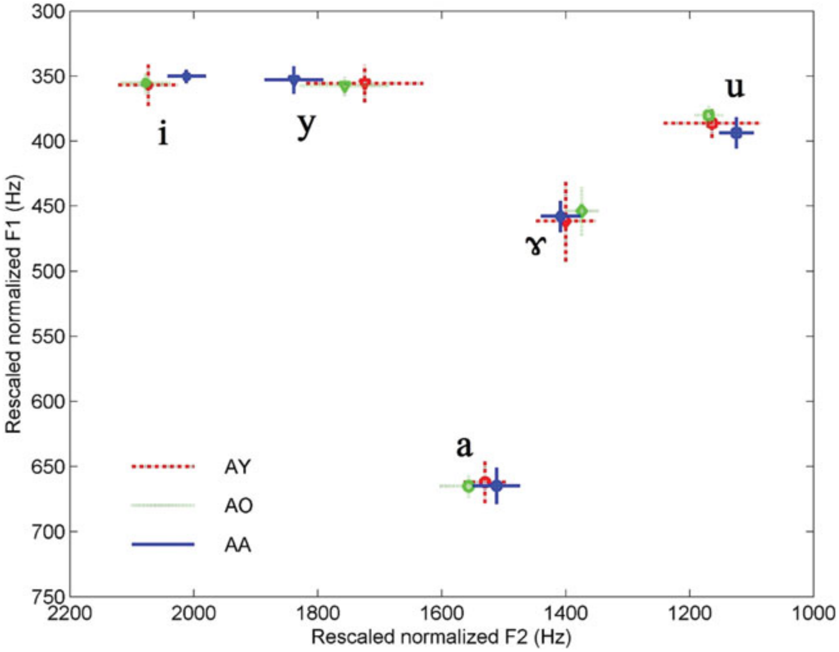

Figure 3 (Colour online) Dispersion of Mandarin vowels for each age group plotted on the basis of the means and standard deviations of rescaled normalized formant frequency values at midpoint location.

examine the effect of vowel and age on F1 and F2 separately. For F1, only the main effect of vowel was significant ($F(4,152) = 2930.303$, $p < .0001$, $\eta_p^2 = .987$). For F2, there was a significant main effect of vowel ($F(4,152) = 1635.322$, $p < .0001$, $\eta_p^2 = .977$) and a vowel by age interaction ($F(4,152) = 6.613$, $p < .0001$, $\eta_p^2 = .258$). The main effect of age just missed significance ($F(2,38) = 3.139$, $p = .055$, $\eta_p^2 = .142$). The significant main effect of vowel on F1 and F2 was anticipated and was not of particular interest in the present study. The significant interaction effect indicated that the three age groups of speakers differed on F2 for certain vowels. The greater standard deviations in children, especially the younger children, suggested that these children exhibited greater variability than the adult speakers in F2 values. This observation, together with the widely scattered vowel tokens in the acoustic space in children relative to adults (shown in Figure 2), indicated that young children were less stable in maintaining articulatory gestures to produce vowels along the tongue advancement scale.

3.3 Vowel space area

The working vowel space is defined by point vowels which are assumed to represent the most peripheral positions of the acoustic vowel space. The size (area) of the vowel working space indexes the developmental change of children's vowel production as a function of increases in vocal tract length (Voperian & Kent 2007). It is also used to examine the nature

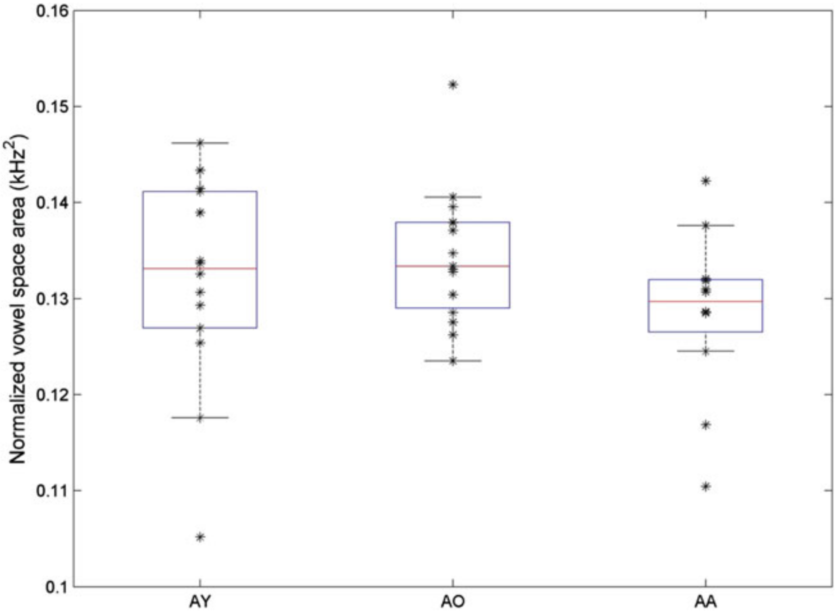

Figure 4 (Colour online) Box plot showing Mandarin vowel space area for each age group. Each box depicts the lower quartile, median and upper quartile. Each asterisk represents a data point of one speaker.

of vowel structure differences across languages and dialects (Chung et al. 2012 and Fox & Jacewicz 2010, respectively). In addition, vowel working space is closely associated with speech intelligibility (Higgins & Hodge 2001, Weismer et al. 2001, Liu, Cao & Kuhl 2005). These studies, collectively, indicate that vowel working space area may represent an important characteristic of vowel production.

The most common approach to measure vowel space area uses the midpoint F1 and F2 values of corner vowels which are defined as /a i u/ in Mandarin. These points serve as the boundary of the triangular vowel space, whose area represents the working vowel space area. In the present study, the normalized vowel space area was calculated on the basis of rescaled normalized formant frequency values shown in Figure 4. We observed the difference in terms of interquartile range which decreased as a function of age. This observation suggested greater individual differences in the child speakers, especially the younger children relative to the adults. A one-way ANOVA result revealed no significant difference among these three groups of speakers in terms of the normalized vowel space area. This result revealed that Mandarin children at this age range, as a group, had developed an adult-like basic vowel framework.

3.4 Formant dynamics

3.4.1 Formant movement pattern

The spectral dynamics patterns for these three age groups are shown in Figure 5. For adult speakers, the five Mandarin vowels demonstrated distinctive formant trajectories. In particular, for the two high front vowels, the vowel /i/ remained relatively stable in both F1 and F2; this resulted in the least formant movement compared to the other vowels. The vowel /y/ showed moderate increase in F1 and decrease in F2. For the low vowel /a/, F2 remained fairly stable while F1 increased at the onset and decreased at the offset. For the high back vowel /u/, F1 did not show much change while F2 showed substantial increase from the 50% to the 80% point. The vowel /ɤ/ demonstrated substantial increase in F1 but relatively no change in F2.

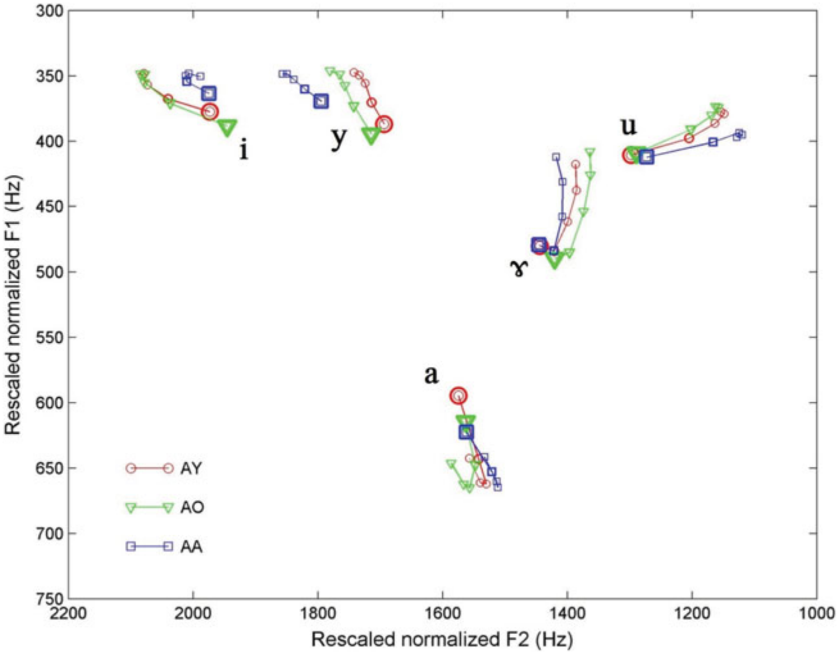

Figure 5 (Colour online) Vowel spectral change plotted on the basis of formant frequencies at five time locations (the 20–35–50–65–80% points) over the course of vowel duration for five Mandarin vowels in the three age groups. The larger size symbol represents the 80% point.

The two groups of children showed comparable formant movement patterns to the adults for the vowels /a u ɤ/. However, the formant trajectories of the vowels /i/ and /y/ produced by the children showed greater movement than those produced by the adults. In addition, the children produced /i/ with a more fronted articulation and /y/ with a more retracted articulation than the adults. Therefore, these two vowels in the adult speakers showed greater positional proximity than those in the children. For the other two vowels /ɤ/ and /u/, the children produced a more fronted /ɤ/ and a slightly lowered /u/ in comparison to the adults.

3.4.2 Trajectory length

The magnitude of formant movement also differed across the three age groups. To better understand the amount of spectral change for each Mandarin vowel, trajectory lengths were calculated on the basis of rescaled normalized formant frequency values over the course of vowel duration (shown in Figure 6). For adult speakers, among these five Mandarin vowels, /u/ had the longest trajectory length, while /i/ had the shortest. In general, the adults produced shorter TLs and showed less variation on the TLs than the children for the vowels /a i y ɤ/ but not for /u/. A two-way repeated-measures ANOVA revealed significant main effects of vowel ($F(4{,}152) = 13.208$, $p < .0001$, $\eta_p^2 = .258$) and age ($F(2{,}38) = 4.876$, $p = .013$, $\eta_p^2 = .204$) as well as a significant vowel by age interaction ($F(8{,}152) = 3.915$, $p = .001$, $\eta_p^2 = .171$). The effect of vowel quality on trajectory length was anticipated and was not of particular interest in the present study. The Bonferroni post hoc test showed that both younger (AY) and older children (AO) produced significantly longer vowel trajectory length than the adults (AA) did. However, no significant difference was found between the two groups of children.

The distribution curves of trajectory lengths for the three age groups for each Mandarin vowel were also compared. These curves were generated using the kernel density estimation function in MATLAB and are shown in Figure 7. In general, kernel density estimation

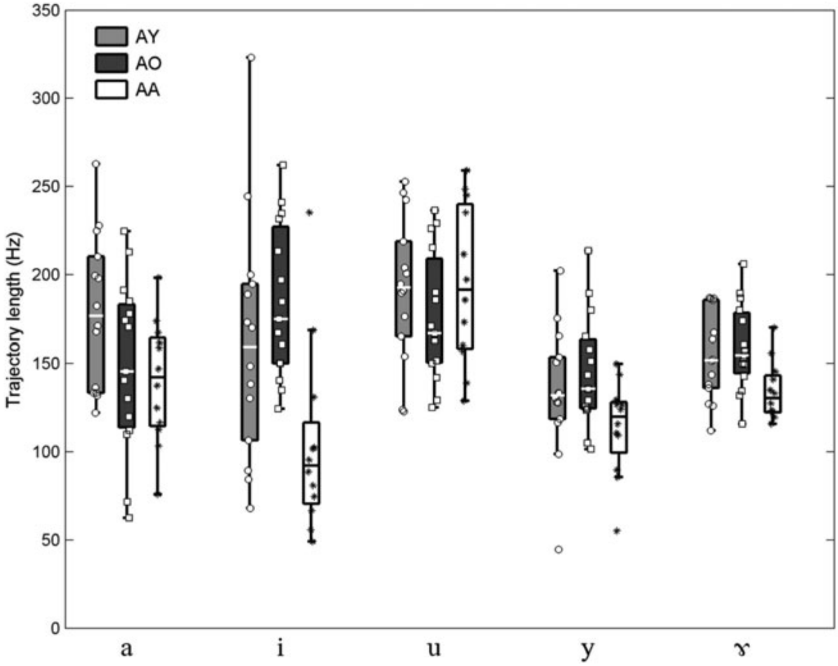

Figure 6 Box plot showing the trajectory length (TL) for each Mandarin vowel in each age group. The dots in the boxes represent the trajectory length of individual speakers within each group. Each box shows horizontal lines at the lower quartile, median, and upper quartile values. The whiskers show the range of the data. The data points out of the whiskers are outliers.

generates a smooth probability density function on the basis of a finite data sample to refer to the distribution of the unknown target population. The kernel density function has two important parameters: the KERNEL FUNCTION and BANDWIDTH. The former determines the shape of the function and the latter determines the extent of smoothness of the function which provides a good estimate of the population density function.

In each distribution curve, the location of the peak indicates where the highest density of the trajectory length is encountered and the spread indicates the variability of the distribution. As shown in Figure 7, the trajectory lengths of the vowels /i y/ in the children were distributed with greater variability and displayed a slight mismatch in terms of peak location from those in the adults. The deviations of children's distribution curves from adult targets again suggested that children at this age had not developed adult-like patterns of formant movement for these two vowels. The distribution curves of the vowel /ɤ/ and /a/ showed gradual change from the younger children to adults, which indicated the approximation of children's trajectory lengths to adults' target. However, for the vowel /u/, the younger children showed highly similar distribution curve to the adults. The curve of the older children even showed less variability than that of adults. This indicated that the children at this age had developed an adult-like spectral dynamic feature in terms of the magnitude of the formant movement for this vowel.

It should be noted that not every group of speakers had a normal distribution curve for every vowel. As shown in Table 3, the distribution curves were skewed for the vowel /a/ of younger children, /i/ of all three groups of speakers and /y/ of older children. The formant trajectories of certain vowels such as /a/ produced by the younger children, /i/ produced by the older children and the adults demonstrated bimodal distributions. This information is not evident from the means and standard deviations. Thus the density function analyses provide important evidence regarding the developmental change of the trajectory length which may not be appropriately shown by ANOVAs.

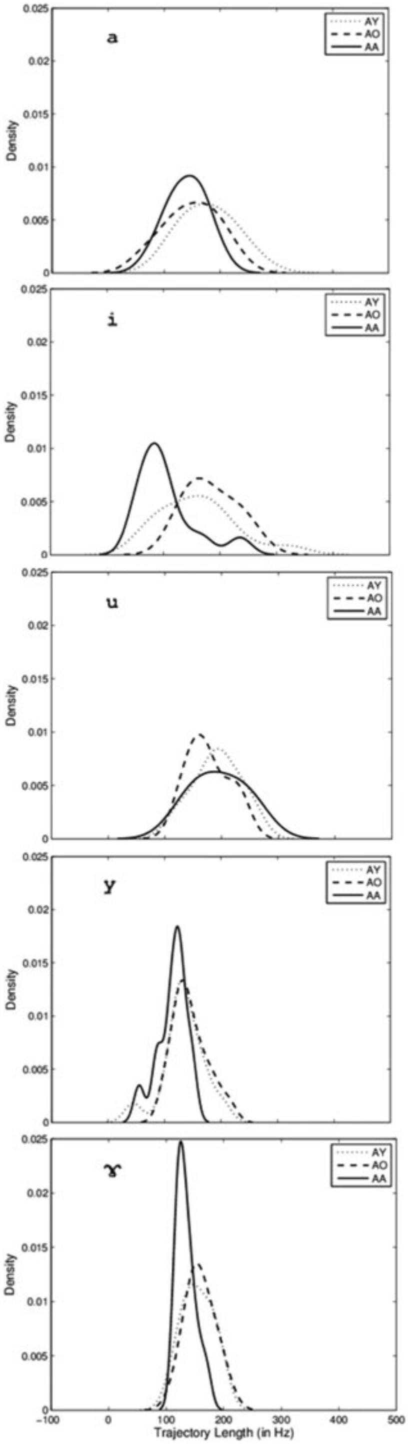

Figure 7 Probability density function of trajectory length for each Mandarin vowel in each age group.

Table 3 Skewness of the distribution curves of trajectory length for individual Mandarin vowels in each age group.

Age group	a	i	u	y	ɤ
AY	0.850	0.973	0.537	0.368	0.461
AO	0.234	0.871	0.413	1.131	0.322
AA	0.217	1.737	0.560	0.150	0.025

4 Summary and discussion

The main purpose of this study was to document the development of the acoustic properties of vowel production in Mandarin-speaking children and to determine to what extent these acoustic features are similar to or different from adults' norms when the effect of vocal tract size is eliminated. A number of acoustic measurements were derived on the basis of the rescaled normalized F1 and F2 values at five temporal locations during the course of vowel duration. The comparison of basic and dynamic vowel acoustic characteristics between the children and adults showed that all five Mandarin monophthongs produced by children deviated from the adult norms in one or more acoustic measures. Although the vowel trajectory lengths of the children's /u/ and /a/ vowels were close to the adults' form, the acoustic features of the three corner vowels did not consistently show greater similarity to the adult targets than did the other non-corner vowels. This is inconsistent with our predictions based on previous transcription studies of vowel acquisition. From this perspective, the present study brings new insights to our understanding of language acquisition, and argues for the importance of acoustic analysis in lieu of studies that use perceptual judgments of vowel quality. Findings from impressionistic studies of vowel development suggest that vowels are fully acquired before three years of age. Although the scatter plots of the midpoint F1 and F2 for the five basic Mandarin vowels in each age group showed clear separation in the vowel space, the size of all five vowel clusters decreased as a function of age. In addition, the children produced all five vowels with longer duration than the adults. They also showed positional deviation from the adult norms. These observations provided evidence to show that children in this age range still show developmental changes in the fine acoustic features of their vowel production.

Shi & Wen (2007) reported that the vowels /i/ and /ɤ/ were acquired earlier than the vowel /u/. However, our acoustic data showed that the acoustic measures of /u/ in children were more similar to the adult norms in comparison to the vowels /i/ or /ɤ/. Compared to the transcription studies of English-speaking children, our acoustic data provided partial evidence to support the early acquisition of corner vowels. For example, in the present study, the vowel /a/ in children was highly similar to that of the adults in all tested acoustic measures. The vowel /u/ produced by the children was also similar to adult norms in many aspects of the acoustic characteristics. However, the high front vowel /i/ in children showed apparent deviation from the adult norms in most acoustic measures. One possible factor related to the acoustic deviation of the vowel /i/ may be that it occurs in significantly fewer consonant contexts than the vowel /a/ due to the phonotactic constraints in Mandarin Chinese. Children have less opportunity to practice this vowel in comparison to the vowel /a/. Therefore, they are less likely to have established early the adult-like target for the vowel /i/. The early acquisition of corner vowels in both Mandarin and English supports the universality of vowel acquisition in linguistically different children. The discrepancies between our results and findings in transcription studies indicates that the acoustic development of speech sounds should be more fully addressed, in addition to traditional phonetic transcriptions in order to better understand the process of speech development in children.

To better understand the nature of vowel production in Mandarin children and adults, we examined vowel dynamic spectral change in both groups. The spectral change of the five Mandarin vowels for adults demonstrated that each Mandarin vowel was characterized by a distinctive formant movement pattern in terms of the shape and movement direction of the formant trajectories. Mandarin-speaking children generally demonstrated compatible patterns of formant movement patterns to Mandarin-speaking adults. Jacewicz et al. (2011b) reported that English-speaking children from different dialect regions had comparable patterns of vowel spectral change as adults in their own dialects. Assmann et al. (2013) found that children as young as five years old exhibited consistent vowel spectral change patterns which were similar to those of adults. In agreement with these findings, the present study revealed that children as young as three years old have established the formant trajectories in a manner similar to adults. However, it can also be observed that certain vowels showed positional change in children relative to adults. Further, the magnitudes of formant movement in certain vowels were also different between children and adults. These observations evidenced the continuing development of these vowels from childhood to adulthood.

On the basis of formant frequency values at the five time locations, we calculated the trajectory length, an evaluation of the magnitude of formant movement in F1 by F2 vowel space over the course of vowel duration. The child speakers generally produced longer trajectory length and showed relatively larger variation than the adults. This finding suggested that the productions of these monophthongal vowels in children were less stable than those in adults. The different patterns of the density functions of trajectory length in certain vowels between children and adults also evidenced the continuing development of children's trajectory lengths. However, one can also notice that some vowels, such as /a/ and /u/, demonstrated similar density function of trajectory length in children and adults. This indicated that the development of vowel spectral change was not parallel across all vowels.

The differences in vowel durations and trajectory lengths between children and adults may indicate the continuing developing speech motor control in children. Previous literature suggested that children moved the articulators more slowly than the adults (Nittrouer 1993) and exhibited much greater variability in the articulatory movement than adults (A. Smith 2006) even though they could move the articulators in a manner similar to adults. Some other studies demonstrated that children went through a nonlinear process in the development of speech motor coordination and exhibited a nonuniform pattern of speech motor development across different articulators (e.g. Sharkey & Folkins 1985, Green, Moore & Reilly 2002, Walsh & A. Smith 2002, A. Smith & Zelaznik 2004). It was found that children who had not developed adult-like abilities in articulatory gesture planning may show a reduction in articulatory movement pace (Kent & Forner 1980, B. L. Smith & Gartenberg 1984). According to these studies, children aged between three and six years may not be able to maintain the articulatory gestures as stable as adults or move the articulators as fast as adults during the process of vowel production due to the continuing development of speech motor skills. The speech motor constraints in children may cause longer segment durations, greater magnitude of formant trajectories and larger variations in acoustic measures in their vowel production.

Among the three groups of speakers, the adults tended to show less variability than the two groups of children (as shown in Figures 2, 3, 4, and 6). For example, the vowel ellipse area displayed a sharp decrease from the younger children to the older children and then to the adults. This pattern suggested that the children showed greater variability than the adults and the variability decreased as a function of age. This finding conforms to the general rule of speech development. However, one can also notice that the younger children approximated the adults to a greater extent on certain acoustic measures than the older children to the adults. For example, the older children, rather than the younger children, produced the vowels /u/ and /ɤ/ with much longer duration than the adults (see Figure 1). The younger children exhibited greater positional approximation to the adults for the vowels /a/ and /ɤ/ than the older children to the adults (see Figure 3). These observations demonstrated that acoustic development is a nonlinear process and the older children may 'overshoot' certain vowels during the process of acoustic development. In addition, the different effect sizes of statistical results also caught our attention when we were interpreting the statistical results of each acoustic measure. In particular, for all repeated-measures ANOVAs, the main effect of vowel demonstrated much greater effect size than the other factors, which indicates that vowel quality accounted for the majority of the variance in these acoustic measures. For those acoustic measures which showed significant main effects of age, smaller effect sizes may be partially associated with large cross-speaker variation which undermined the explanatory power of this factor.

5 Conclusion

In general, the comprehensive examination of vowel acoustic characteristics in three-to-six-year-old Mandarin-speaking children revealed that children in this age range still show considerable acoustic differences from adults, which indicates that the acoustic development of children's vowel production is a long-term process. However, though meticulous and comprehensive, the present study was based on a relatively small number of subjects in each age group. In future studies, more subjects of both sexes should be recorded from a wider age range to examine whether the factor of sex causes discernable differences in children's vowel acoustic characteristics when the effect of vocal tract size is excluded. In addition, a better experiment design with more strictly controlled speech materials and an alternative recording paradigm of spontaneous speech elicitation should be used.

Acknowledgements

The authors would like to thank the editor, Professor Amalia Arvaniti, and three anonymous reviewers for their valuable and insightful comments and suggestions to improve the quality of the paper. The authors would also like to thank the children and their parents for their participation in this study.

References

Adank, Patti, Roel Smits & Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116, 3099–3107.

Assmann, Peter F., Terrance M. Nearey & Sneha V. Bharadwaj. 2013. Development patterns in children's speech: Patterns of spectral change in vowels. In Geoffrey Stewart Morrison & Peter F. Assmann (eds.), *Vowel inherent spectral change*, 199–230. Berlin: Springer.

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

Busby, Peter A. & G. L. Plant. 1995. Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America* 97, 2603–2606.

Chang, Chiung-Yun. 2010. *Dialect differences in the production and perception of Mandarin Chinese tones*. Ph.D. dissertation. The Ohio State University.

Chung, Hyunju, Eun Jong Kong, Jan Edwards, Gary Weismer, Marios Fourakis & Youngdeok Hwang. 2012. Cross-linguistic studies of children's and adults' vowel spaces. *The Journal of the Acoustical Society of America* 131, 442–454.

Davis, Barbara L. & Peter F. MacNeilage. 1990. Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research* 33, 16–27.

Edwards, Jan & Mary E. Beckman. 2008. Methodological questions in studying consonant acquisition. *Clinical Linguistics and Phonetics* 22, 937–956.

Fox, Robert A. & Ewa Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America* 126, 2603–2618.

Fox, Robert A. & Ewa Jacewicz. 2010. Dialect and generational differences in vowel space areas. In Antonis Botinis (ed.), *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 37–40. ISCA and the University of Athens, Greece.

Gilbert, Harvey R., Michael P. Robb & Yang Chen. 1997. Formant frequency development: 15 to 36 months. *Journal of Voice* 11(3), 260–266.

Green, Jordan R., Christopher A. Moore & Kevin J. Reilly. 2002. The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research* 45, 66–79.

Hare, Gayle J. 1983. Development at 2 years. In Irwin & Wong (eds.), 55–88.

Harrington, Jonathan & Stephen Cassidy. 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech* 37, 357–373.

Higgins, Carolyn M. & Megan M. Hodge. 2001. F2/F1 vowel quadrilateral area in young children with and without dysarthria. *Canadian Acoustics* 29, 66–68.

Hillenbrand, James, Laura A. Getty, Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97(5), 3099–3111.

Hoole, Philip & Fang Hu. 2004. Tone–vowel interaction in standard Chinese. Presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages.

Irwin, John V. & Seok P. Wong (eds.). 1983. *Phonological development in children: 18–72 months*. Carbondale, IL: Southern Illinois University Press.

Jacewicz, Ewa, Robert A. Fox & Joseph Salmons. 2011a. Cross-generational vowel change in American English. *Language Variation and Change* 23(1), 45–86.

Jacewicz, Ewa, Robert A. Fox & Joseph Salmons. 2011b. Regional dialect variation in the vowel systems of typically developing children. *Journal of Speech, Language, and Hearing Research* 54, 448–470.

Jacewicz, Ewa, Robert A. Fox & Joseph Salmons. 2011c. Vowel change across three age groups of speakers in three regional varieties of American English. *Journal of Phonetics* 39, 683–693.

Jacewicz, Ewa, Joseph Salmons & Robert A. Fox. 2009. Prosodic conditioning, vowel dynamics and sound change. In Caroline Féry, Frank Kügler & Ruben van de Vijver (eds.), *Variation and gradience in phonetics and phonology*, 99–124. Berlin: Mouton de Gruyter.

Jeng, Heng-hsiung. 1979. The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity. *Proceedings of 9th International Congress of Phonetic Sciences* (ICPhS IX), vol. 2, 155–161. Stockholm: Institute of Phonetics.

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

Kent, Raymond D. & Ann D. Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America* 72(2), 353–365.

Kent, Raymond D. & Charles Read. 1992. *The acoustic analysis of speech*. San Diego, CA: Singular Publishing.

Kuhl, Patricia K. & Andrew N. Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and development change. *The Journal of the Acoustical Society of America* 100(4), 2425–2438.

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

Lee-Kim, Sang-Im. 2014. Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association* 44(3), 261–282.

Lin, Maocan. 1988. Putonghua shengdiao de shengxue texing he zhijue zhengzhao [The acoustic characteristics and perceptual cues of tones in Standard Chinese]. *Zhongguo Yuwen* [Chinese Language and Writing] 204, 182–193.

Liu, Huei-Mei., Feng-Ming Cao & Patricia L. Kuhl. 2005. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America* 117, 3879–3889.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *The Journal of the Acoustical Society of America* 49, 606–608.

Milenkovic, Paul. 2003. TF32 software program. University of Wisconsin–Madison.

Nearey, Terrance M. 2013. Vowel inherent spectral change in the vowels of North American English. In Geoffrey Stewart Morrison & Peter F. Assmann (eds.), *Vowel inherent spectral change*. Berlin: Springer.

Nearey, Terrance M. & Peter F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America* 80, 1297–1308.

Nittrouer, Susan. 1993. The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research* 36, 959–972.

Otomo, Kiyoshi & Carol Stoel-Gammon. 1992. The acquisition of unrounded vowels in English. *Journal of Speech and Hearing Research* 35, 604–616.

Paschall, Leslie. 1983. Development at 18 months. In Irwin & Wong (eds.), 27–54.

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

Shi, Feng & Baoying Wen. 2007. Hanyu Putonghua ertong de yuanyin fazhan [Vowel development of Mandarin-speaking children]. *Zhongguo Yuwen* [Chinese language and Writing] 5, 444–454.

Si, Yuying. 2006. Putonghua ertong yuyin xide de ge'an yanjiug [The acquisition of Puthonghua Phonology: A case study]. *Dangdai Yuyanxue* [Contemporary Linguistics] 1, 93–94.

Smith, Anne. 2006. Speech motor development: Integrating muscles, movements and linguistic units. *Journal of Communication Disorders* 39, 331–349.

Smith, Anne & Howard N. Zelaznik. 2004. The development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology* 45, 22–33.

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

Smith, Bruce L. 1991. Relationships between duration and temporal variability in children's speech. *The Journal of the Acoustical Society of America* 91, 2165–2174.

Smith, Bruce L. & Terri E. Gartenberg. 1984. Initial observations concerning developmental characteristics of labio-mandibular kinematics. *The Journal of the Acoustical Society of America* 75(5), 1599–1605.

Smith, Neilson Voyne. 1973. *The acquisition of phonology*. New York: Cambridge University Press.

Stoel-Gammon, Carol & Paula Beckett Herrington. 1990. Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics & Phonetics* 4(2), 145–160.

Templin, Mildred C. 1957. Certain language skills in children: Their development and interrelationships (Institute of Child Welfare Monographs 26). Minneapolis, MN: University of Minnesota Press.

Thomas, Erik R. & Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. http://ncslaap.lib.ncsu.edu/tools/norm/about_norm1.php#scaling, accessed on 9 March 2013.

Vihman, Marilyn May. 1993. Variable paths to early word production. *Journal of Phonetics* 21, 61–82.

Vihman, Marilyn May, Charles A. Ferguson & Mary Elbert. 1986. Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics* 7, 3–40.

Vorperian, Houri K. & Raymond D. Kent. 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research* 50(6), 1510–1545.

Walsh, Bridget & Anne Smith. 2002. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research* 45, 1119–1133.

Waterson, Natalie. 1971. Child phonology: A prosodic view. *Journal of Linguistics* 7, 179–211.

Weismer, Gary, J-Y. Jeng, Jacqueline S. Laures, Raymond D. Kent & Jane F. Kent. 2001. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica* 53, 1–18.

Wellman, Beth L., Ida Mae Case, Ida Gaarder Mengert & Dorothy Edith Bradbury. 1931. Speech sounds of young children, In G. D. Stoddard (ed.), *University of Iowa Studies in Child Welfare* 5(2), 7–80. Iowa City, IO: University of Iowa Press.

Xu, Li., Yuhjung Tsai & Bryan E. Pfingst. 2002. Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *The Journal of the Acoustical Society of America* 112(1), 247–258.

Yuan, Jiahong. 2013. The spectral dynamics of vowels in Mandarin Chinese. *Proceedings of the 14th Annual Conference of the International Speech Communication Association* (INTERSPEECH 2013), 1193–1197.

Zhu, Hua & Barbara Dodd. 2000. The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* 27(1), 3–42.

Zhou, Ning & Li Xu. 2008. Lexical tone recognition with spectrally mismatched envelopes. *Hearing Research* 246, 36–43.