



Acoustic development of vowel production in American English children

Jing Yang, Robert A. Fox

Department of Speech and Hearing Science, The Ohio State University, Columbus, OH, USA

yang.1198@osu.edu, fox.2@osu.edu

Abstract

Children's vowel acquisition has long been examined on the basis of transcription-based evaluations of the accuracy rate of the vowel production in children before 5 years of age. This study examines the development of static and dynamic acoustic features in children between 3 and 7 years of age by comparing the acoustic features of children with those of adults. All acoustic analyses were based on the normalized formant frequency values to exclude the effect of different vocal tract size. The increasing compactness of individual vowel categories in the acoustic space evidenced the refinement of phonetic features in children in this age range. In addition, the vowel dispersion pattern of certain vowel plotted on the basis of formant frequency values at 5 temporal locations demonstrated positional change as well as differences in terms of the trajectory length. Results demonstrate that the acoustical development of vowels from children to adult norms is likely a long-term, graduate but not necessarily continuous process.

Index Terms: speech development, monolingual English, vowel acoustics

1. Introduction

Compared to many other languages in the world, English has a relatively large vowel inventory. How English children acquire the vowel contrasts and establish the vowel system in their native language has long been of research interest. In general, there are two different approaches to the examination of children's vowel development. The first one involves the evaluation of the accuracy rate of children's vowel production on the basis of phonetic transcription by a native adult speaker/phonetician to determine the age of acquisition of individual speech sounds [1, 2, 3]. However, this subjective evaluation metric can be affected by a number of factors including the child's age, the child's physical and personality characteristics, the intelligibility of the child's speech, the linguistic context in which the vowel was produced, etc. Transcribers can even be biased by successive judgments of produced phonetic segments [4]. Therefore, even if a sound is described as having been "acquired" by a child at a particular age does not mean that it is fully (and finally) developed. Some studies argued that children typically do not master adult-like features of speech until 8-years or even later [5, 6]. Indeed, more objective analysis is needed to capture the subtle phonetic differences which can not be demonstrated even by narrow transcription. The documentation of phonetic changes from child- to adult-like targets can not only further our understanding in the nature of speech development, but also shed light on the development of underlying cognitive process.

Acoustic analysis (rather than phonetic transcription) has also been widely used to document the developmental change of acoustic features in childhood speech as a result of the lengthening of the vocal tract. Collectively, many previous works [7, 8, 9] have shown a pattern of decreasing formant

frequency values and vowel duration as a function of chronological age. Except for the decreasing formant frequency values resulting from the lengthening of the vocal tract, children's speech production also show other developmental changes resulting from the maturation of articulators, improved motor control, growth of vocabulary size and increased cognitive abilities. A large body of research [10, 11, 12] has revealed that a child's articulators go through nonlinear and nonuniform developmental patterns. The protracted development of speech motor control is manifested in both the temporal and spectral properties of children's speech production [13]. In addition to the maturation of speech motor control, researchers have also found that the child's growth in vocabulary size affects her sensitivity to phonotactic probability which then influences the accuracy of her speech production [14].

Thus, although children can articulate perceptibly accurate and clear speech sounds at quite early age, their speech production may still show inconsistencies and differences relative to adulthood speech. The present study aims to expand previous literature on the developmental changes on the acoustic characteristics of vowel production. Of particular interest is the acoustic development of both static and dynamic vowel features in early aged children and the extent to which these features are similar to or different from those of adults.

Formant frequencies at the midpoint location are often used to represent the acoustic properties of vowels, assuming it is basically a steady-state signal. However, there is much evidence that vowels are also characterized by the inherent dynamic spectral changes [15, 16]. Vowel spectral change also conveys important information associated with speaker's dialect [17] and speaker's generation [18]. In addition, vowel spectral change is also related to speakers' speech motor control because during the process of vowel production, even slight movement of articulators may change the resonance feature of the vocal tract, which may cause measurable change of formant frequency values. Therefore, the dynamic vowel spectral change resulting from the ongoing coordination of articulators is of importance to the development of children's speech motor control.

Given the significance of vowel dynamic features, more attention has been recently been paid to the development of vowel dynamic features in children [19, 20]. In general, these studies have found that children have comparable patterns of vowel spectral change as adults. However, the majority of the available studies have addressed vowel dynamic features in relatively older children. Relatively few of them examined the nature of formant dynamics in younger children. In addition, very little research has been done in quantitatively defining children's vowel spectral change as has been done with adults [17]. To date, it remains unclear whether children at a younger age can produce speech sounds with adult-like dynamic acoustic features. This study seeks to examine this question while controlling for physical maturation of these children's

vocal tracts (through the use of standard vowel normalization schemes).

2. Method

2.1. Speakers

Fifteen native English speaking children (7 girls and 8 boys) aged 3 to 7 years and six native English speaking adults (6 females) aged 30 to 44 years participated in this study. The speakers were divided into 3 age groups (**AY**: 3-5, 6 children; **AO**: 5-7, 9 children; **AA**: adults). All children were born and raised in central Ohio region (Columbus, Ohio). Likewise, all adults were from central Ohio and currently live in Columbus, Ohio. The six female adult speakers were mothers of the child speakers in the present study. All speakers were reported having no speech and language disorders.

2.2. Stimuli

The recording material included 20 English monosyllabic/disyllabic words (see Appendix for the word list) containing 10 American English vowels /i ɪ e ε æ u o ɑ ʌ/ (due to the merger of /ɔ/ and /ɑ/ in most dialects of American English including Ohio English, just one vowel /ɑ/ was selected in the present study). The consonants preceding the target vowels were stop consonants except for /f/ in “feet”. The consonants following the target vowels were voiceless stops or fricatives. The selection of stimulus words was on the basis of familiarity, word frequency and picturability (as stimulus targets were presented through pictures rather than written words).

2.3. Procedures

The speech samples were collected through a word-repetition task under the control of a custom MATLAB program. The word-repetition task instead of spontaneous speech elicitation was used because the former method enables experimenters to have better control of the stimulus presentation and to ensure the speakers produce the specific target words as expected [21]. The word-repetition experiment included two blocks of recording. In each randomly ordered block (the same random order was used for both recording blocks) pictures representing target words were presented on a computer monitor. Each speaker was seated in front of a laptop computer in a quiet room and repeated each target word immediately following an audio prompt produced by a native English speaker. A Shure SM10A head-mounted microphone was situated approximately 1-inch from the speakers’ mouth and all productions were recorded and digitized directly onto a hard drive disk at a 16-bit quantization rate and 44.1 kHz sampling rate.

2.4. Acoustic measurements

2.4.1. Formant frequencies

Prior to acoustic analysis, all tokens were down-sampled to 11 kHz. To provide an estimate of formant dynamics (and reduce the redundant calculation caused by more dense sampling), formant frequencies were measured at five equidistant temporal locations (the 20-35-50-65-80% points) using spectrographic analysis program TF32 [22]. The onsets and offsets of the vowel were excluded to eliminate the effect of consonant context (i.e., context-dependent formant transitions). The landmark locations of vowel onset and offset were located

by hand on the basis of the waveform as well as visual checks of a spectrogram following standard criteria [23]. For example, the location of vowel onset was defined at the point of the initial zero crossing at which the amplitude of the waveform increased significantly following stop closure release or cessation of frication. Vowel offset was set at the zero crossing point where the amplitude of the waveform dropped significantly due to the following stop closure or frication.

2.4.2. Vowel normalization

A set of normalized formant frequency values was then generated using the method in Lobanov (1971) [24] (found to be one of the most effective normalization methods [25]) to eliminate the effect of vocal tract length different among speakers as a function of chronological age (which, in turn, affects raw formant frequency values). The normalized formant frequency values were then rescaled into Hertz-like values using the method described in Thomas and Kendall (2007) [26] to facilitate interpretation of the normalized vowel spaces.

2.4.3. Trajectory length

Trajectory lengths (TL), 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%), were calculated on the basis of the rescaled normalized F1 and F2 values in the acoustic vowel space [17] using the following formula:

$$TL = \sum_{n=1}^4 VSL_n \quad (1)$$

where the length of each vowel section (VSL) is calculated based on the formula:

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

TL provides an unsigned measure of the magnitude of vowel movement in the F1 x F2 acoustic plane over the course of vowel duration between the 20 to 80% points.

3. Results

3.1. Midpoint F1 by F2 vowel space

Scatter plots of vowels in the F1 x F2 plane using rescaled normalized F1 and F2 at midpoint locations are shown in Figure 1. In each panel, ellipses are drawn which enclose 95% of the samples in each vowel category. The areas enclosed by these ellipses for each age group (a measure of the dispersion of recorded vowel exemplars for each category) are shown in Table 1. As can be seen, there are no clearly observable changes between younger and older children in terms of the positional distribution of the front vowel categories. However, the back vowels show significant observable changes particularly in the vowel /u ʊ o/. For adult speakers, all 10 vowel categories are well separated and tightly concentrated. The biggest change from children to adults happened on the complete separation of high front vowels /i ɪ e/ and high back vowels /u ʊ/. In addition, the size of the ellipse for each acoustic vowel category shows significant reduction in adults compared to the children. The decreasing size of the area of the ellipses implies the continuing refinement of vowel acoustic features as a function of maturation. It should be

pointed out that although the vowel /u/ showed a decreasing ellipse size from children to adults, compared to other vowels, this vowel still displayed relatively large variation along the F2 axis even in adult speakers. That is, the adult speakers varied in the frontness of the vowel /u/. As mentioned earlier, all native English speakers were from central Ohio and one of the important characteristics of this dialect of American English is the tendency to front /u/. The speakers in the present study indeed presented the pattern of sound change in the vowel /u/ occurring in the Midland dialect.

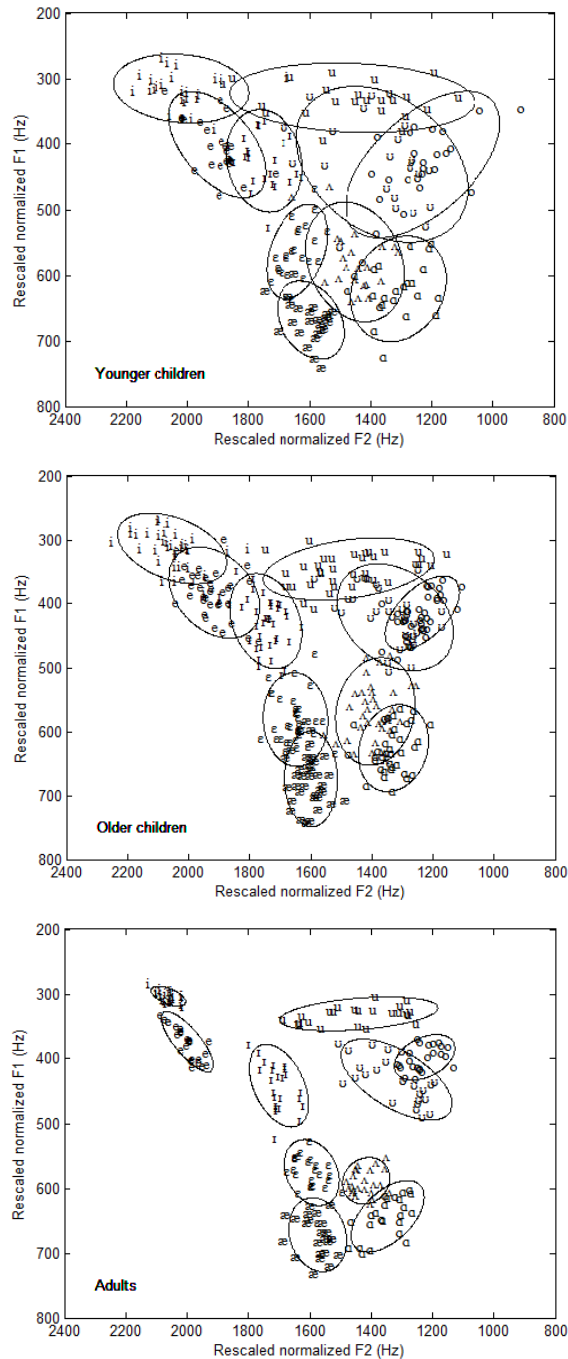


Figure 1: Scatter plot of vowels on the basis of midpoint formant frequencies for individual English vowel categories in each age group.

Table 1. The area of ellipse of each English vowel category in F1 x F2 vowel space across three age groups (in Hz²)

	AY	AO	AA
i	36,272	28,421	2,210
ɪ	30,112	26,676	17,475
e	34,160	31,521	7,665
ε	20,852	24,386	13,893
æ	19,451	20,438	16,658
u	66,214	40,415	19,723
ʊ	99,234	47,789	28,244
o	72,639	18,286	9,994
ɑ	38,874	24,257	17,397
ʌ	45,015	34,222	8,662

3.2. Formant movement pattern

Figure 2 shows the formant trajectories (in the F1 x F2 plane) for these 10 English vowels in each age group. For adult speakers, the 10 English vowels display distinctive vowel formant frequency dynamic patterns in both the position and trajectory of the formant tracks. In children, the general formant trajectories in each vowel to a large extent are very similar to that of adults. However, positional changes and specific differences in the patterns of formant tracks can still be observed in certain vowels. For example, /e/ and /o/ were produced in a more fronted position as a function of speaker age. The vowel /ɪ/ showed a downward movement from children to adults and /ɑ/ demonstrated a forward and slightly downward movement. The vowel /ʌ/ also showed position change from children to adults. In addition, the formant trajectories for /ʌ/ produced by children were less curved than that produced by adults.

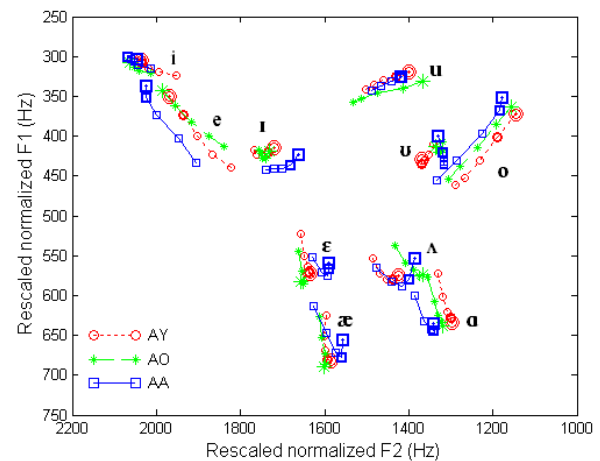


Figure 2: Vowel dispersion of spectral change plotted on the basis of formant frequencies at five time locations (20-35-50-65-80%) over the course of vowel duration for 10 English vowels across three age groups. The larger size symbol represents the 80% point.

3.3. Trajectory length

Table 2 summarizes the mean and standard error of trajectory length on the basis of rescaled normalized formant frequency

values for each vowel in each age group. The trajectory length increased from children to adults for most vowels (/i e ε æ o a ʌ/). But for certain vowels like /i u ʊ/, the trajectory length showed a slight decrease or no apparent change from children to adults. In order to better understand how the age factor influenced the trajectory length for each separate vowel, one-way ANOVA tests were applied and Tukey HSD was used to examine the nature of significant differences between the age groups. The results show that for the vowel /æ ʌ/, adults have significantly longer trajectory lengths than both younger and older children. For the vowel /i a/, adults have significantly longer trajectory length than older children. For the vowel /ε/, one-way ANOVA result shows significant effect of age, but Tukey HSD tests just show marginal significance between adults and both younger and older children.

Table 2. Summary of statistical results from one-way ANOVAs for formant trajectory length for the effect of age on each English vowel (Mean and standard error of trajectory length are in Hz).

	AY M (SE)	AO M (SE)	AA M (SE)	<i>p</i>	η^2	Tukey HSD
i	134 (20)	101 (8)	105 (7)	0.147	0.191	
ɪ	91 (12)	74 (2)	113 (13)	0.021	0.35	AO<AA
e	190 (20)	181 (21)	194 (16)	0.898	0.012	
ɛ	98 (9)	101 (4)	132 (15)	0.039	0.303	AY<AA, AO<AA
æ	90 (7)	105 (6)	150 (22)	0.012	0.39	AY<AA, AO<AA
u	134 (20)	199 (37)	137 (19)	0.237	0.148	
ʊ	87 (10)	82 (7)	84 (7)	0.883	0.017	
o	191 (22)	188 (15)	225 (12)	0.249	0.143	
ɑ	125 (8)	101 (5)	135 (11)	0.014	0.377	AO<AA
ʌ	102 (6)	109 (7)	150 (10)	0.002	0.514	AY<AA, AO<AA

4. Summary and discussion

This study examined the development of acoustic properties in vowel production by native English children aged 3 to 7. Of particular interest is how children at this age range are similar to or different from adults in both static and dynamic vowel features. By means of vowel normalization, we eliminated or reduced the effect of vocal tract size on the formant frequency values. The results demonstrated that children still show developmental changes in their organization of acoustic vowel categories and did show some different patterns of vowel dynamics from adults. These developmental changes indicate that children are still refining their phonetic characteristics and approximating adult-like targets after 3 years of age, the timeline of accomplishment of vowel acquisition argued by most previous studies.

As shown in the scatter plots, younger children showed considerable positional overlap in the back vowels, while the front vowels demonstrated less variability. In the older children, the overlap in the back vowels decreased while

positions of the front vowels remained relatively stable. The distribution of the vowels in the younger children is characterized by a higher degree of centralization (than in the older children and the adults) likely a product of infant vocalization patterns [27]. In the older children, the overlaps among the vowel clusters decrease, which indicates establishment of better defined vowel categories.

The apparent reduction of acoustic overlap in back vowels may be associated with the protracted development of children's speech motor control. Previous studies have shown that the coordination of lower lip and upper lip demonstrate a gradual development pattern of maturation. Most of back vowels in English are round vowels which are produced with lips protruded or rounded. Due to the protracted development of motor skills related to the lips, the acoustic manifestation of the back vowels demonstrated greater variability.

In terms of the spectral dynamics, both groups of children, in general, show comparable formant movement patterns as those of English adults. However, the formant tracks produced by children do not match perfectly with those produced by adults in either the relative positions of the vowels nor in the lengths of the formant trajectories. The downward movement of /ɪ/ and clockwise movement of /ɑ/ from children to adults are consistent with the previous finding [28], which provide extra evidence for the mechanism of cross-generational vowel change in Ohio English. Except for the positional change, children's vowel productions also differ from adults' in the trajectory length. Children's vowel productions are characterized by shorter trajectory lengths.

In sum, the detailed comparison of acoustic vowel characteristics between children and adults revealed that children's vowel production undergo substantial developmental change other than the effect of lengthening of vocal tract. The acoustical development of vowels from children to adult norms is a long-term process.

5. Acknowledgements

This research was supported, in part, by the Ruth Becky Irwin Fund in the Department of Speech and Hearing Science at The Ohio State University. We also like to thank the subjects and their parents for participating in this study.

Appendix: Word list used to elicit speech samples

vowel	word
i	feet geese
ɪ	kiss kitchen
e	cake tape
ɛ	desk pet
æ	cat bat
u	boot goose
ʊ	book cookie
o	coat boat
ɑ	box stop
ʌ	cup duck

6. References

- [1] Wellman, B. L., Case, I. M., Mengert, I. B. and Bradbury, D. E., "Speech sounds of young children", In G.D. Stoddard [Ed], University of Iowa studies in child welfare, 5(2). Iowa City, IO: University of Iowa Press, 1931.
- [2] Davis, B. L. and MacNeilage, P. F., "Acquisition of correct vowel production: A quantitative case study", *Journal of Speech and Hearing Research*, 3:16-27, 1990.
- [3] Stoel-Gammon, C. and Beckett Herrington, P., "Vowel systems of normally developing and phonologically disordered children", *Clinical Linguistics & Phonetics*, 4:145-160, 1990.
- [4] Shriberg, L. D. and Kent, R. D., *Clinical Phonetics*, NY, Pearson, 2003.
- [5] Sanders, E. K., "When are speech sounds learned?", *Journal of Speech and Hearing Disorders*, 37:55-63, 1972.
- [6] Tingley, B. M. and Allen, G. D., "Development of speech timing control in children", *Child Development*, 46: 186-194, 1975.
- [7] Vorperian, H. K. and Kent, R. D., "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data", *Journal of Speech, Language, and Hearing Research*, 50 (6):1510-1545, 2007.
- [8] Busby, P. A. and Plant, G. L., "Formant frequency values of vowels produced by preadolescent boys and girls", *Journal of the Acoustical Society of America*, 97(4):2603-2606, 1995.
- [9] Lee, S., Potamianos, A. and Narayanan, S., "Acoustics of children's speech: Developmental changes of temporal and spectral parameters", *Journal of the Acoustical Society of America*, 105(3):1455-1468, 1999.
- [10] Green J. R., Moore C. A. and Reilly K. J., "The sequential development of jaw and lip control for speech", *Journal of Speech, Language, and Hearing Research*, 45: 66-79, 2002.
- [11] Sharkey, S. G. and Folkins, J. W., "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, 1985.
- [12] Walsh, B. and Smith, A., "Articulatory movement in adolescents: evidence for protracted development of speech motor control processes", *Journal of Speech, Language, and Hearing Research*, 45:1119-1133, 2002.
- [13] Nittrouer, S., "The emergence of mature gestural patterns is not uniform: evidence from an acoustic study", *Journal of Speech and Hearing Research*, 36:959-972, 1993.
- [14] Edwards, J., Beckman, M. E. and Munson, B., "The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition", *Journal of Speech, Language and Hearing Research*, 47:421-436, 2004.
- [15] Hillenbrand, J., Getty, L. A., Clark, M. J. and Wheeler, K., "Acoustic characteristics of American English vowels", *Journal of the Acoustical Society of America*, 97(5):3099-3111, 1995.
- [16] Neary, T. M. and Assmann, P. F., "Modeling the role of inherent spectral change in vowel identification", *Journal of the Acoustical Society of America*, 80:1297-1308, 1986.
- [17] Fox, R. A. and Jacewicz, E., "Cross-dialectal variation in formant dynamics of American English vowels", *Journal of the Acoustical Society of America*, 126:2603-2618, 2009.
- [18] Jacewicz, E., Fox, R. A. and Salmons, J., "Cross-generational vowel change in American English", *Language Variation and Change*, 23(1):45-86, 2011.
- [19] Assmann, P. F. and Katz, W. F., "Time-varying spectral change in the vowels of children and adults", *Journal of the Acoustical Society of America*, 108(4):1856-1866, 2000.
- [20] Jacewicz, E., Fox, R. A. and Salmons, J., "Regional dialect variation in the vowel systems of typically developing children", *Journal of Speech, Language and Hearing Research*, 54:448-470, 2011.
- [21] Edwards, J. and Beckman, M. E., "Methodological questions in studying consonant acquisition", *Clinical Linguistics and Phonetics*, 22(12):939-958, 2008.
- [22] Milenkovic, P., TF32 software program. University of Wisconsin, Madison, 2003.
- [23] Kent, R. D. and Read, C., *The acoustic analysis of speech*, San Diego, CA: Singular Publishing, 1992.
- [24] Lobanov, B. M., "Classification of Russian vowels spoken by different listeners", *Journal of the Acoustical Society of America*, 49:606-608, 1971.
- [25] Adank, P., Smits, R. and van Hout, R., "A comparison of vowel normalization procedures for language variation research", *Journal of the Acoustical Society of America*, 116:3099-3107, 2004.
- [26] Thomas, E. R. and Kendall, T., "NORM: The vowel normalization and plotting suite", online: http://ncslaap.lib.ncsu.edu/tools/norm/about_norm1.php#scaling, accessed on 9 Mar 2013.
- [27] Kuhl, P. K. and Meltzoff, A. N., "Infant vocalizations in response to speech: Vocal imitation and developmental change", *Journal of the Acoustical Society of America*, 100 (4):2425-2438, 1996.
- [28] Jacewicz, E., Fox, R. A. and Salmons, J., "Vowel change across three age groups of speakers in three regional varieties of American English", *Journal of Phonetics*, 39: 683-693, 2011.