Acoustic Properties of Vowel Production in Mandarin-English Bilingual and Corresponding Monolingual Children

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Jing Yang

Graduate Program in Speech and Hearing Science

The Ohio State University
2014

Dissertation Committee:

Dr. Robert A. Fox, Advisor
Dr. Rebecca J. McCauley
Dr. Marjorie K. M. Chan
Dr. Cynthia G. Clopper
Abstract

This dissertation includes four separate but interrelated studies which serve to address two main goals: First, studies 1 and 2 investigate the development of acoustic characteristics of vowel production in 3-6 years old monolingual Mandarin and monolingual English children relative to corresponding monolingual adults. Second, studies 3 and 4 examine the phonetic features associated with the process of language separation and L1-L2 interaction in young bilingual Mandarin-English children.

A number of studies used transcription-based accuracy rating method and have shown that vowels are relatively easy to acquire and vowel acquisition is generally completed prior to three years of age. In order to examine whether children exhibit continuing development of acoustic features after three years of age, studies 1 and 2 compared both static and dynamic acoustic characteristics of vowel production between monolingual Mandarin children and adults (study 1) as well as monolingual English children and adults (study 2). Both Mandarin children and English children showed increasing compactness of individual vowel categories in the acoustic space, which evidenced the refinement of acoustic details in children in this age range. In addition, for both Mandarin and English children, the vowel dynamics also provide evidence to show the developmental change for certain vowels. Study 3 documented the emergence of bilingualism in a young boy from the age of 3:7 to 5;3. The child initially utilized his L1 base in building the L2 vowel system. The L1-L2 separation began through a drastic
restructuring of his working vowel space to create maximal contrast between the two languages. This abrupt partitioning was accomplished by temporarily creating a reduced L2 vowel space, which gradually expanded as the child “added” L2 vowels to his L2 system. While the general shape of his L1 vowel space remained unchanged throughout, L1 developmental processes and influence of L2 on L1 were also in effect.

Study 4 examined the extent to which the L1-L2 interaction effect changes the phonetic features of vowel productions of young bilingual Mandarin-English-speaking children. It is found that young bilingual children with low proficiency in English (L2) preserved the basic acoustic features of Mandarin (L1). But their production of English was strongly affected by their Mandarin. In particular, they transferred both static and dynamic vowel features from L1 to L2. Young bilingual children with high proficiency in English produced English vowels in a near-monolingual manner. However, they still show different acoustic features in certain aspects of their vowel production. In addition, due to the influence of English, their production of Mandarin vowels showed systematic differences from monolingual Mandarin children in both static and dynamic vowel features.

In sum, this dissertation provided evidences to show the continuing acoustic development in children after three years of age regardless of their language background. As a result of on-going acoustic development in young children, bilingual children showed an active bi-directional interaction effect between L1 and L2. But the strength and size of the interaction effect is determined by their language experience in L2.
Dedication

Dedicated to my family.
Acknowledgments

I would like to express my deepest appreciation to my advisor, Professor Robert Fox, for his excellent guidance, patience, caring and mentorship provided to me. Over the past six years, his serious scientific attitude, rigorous scholarship, profound knowledge and genial personality deeply infected and inspired me. I feel truly fortunate to have had the chance to study and work with him not just because of the maximum freedom he gave me to explore my own interests but also because of his selfless help whenever I encountered difficulties.

I would like to sincerely thank my committee members: Dr. Rebecca McCauley, Dr. Marjorie Chan and Dr. Cynthia Clopper for their support, understanding, encouragement, insightful advice and constructive comments throughout the process. Without their persistent help, this dissertation would not have been possible.

I would like to express my special gratitude to Dr. Ewa Jacewicz who inspired me to appreciate the joys and challenges of an academic career. I am very grateful to Dr. Michael Trudeau for his fatherly advice and great support.

I am deeply indebted to my family members who always believe in me and encourage me to pursue my dreams. Their unconditional love and support make me powerful and confident. I am also grateful to all of my closest and dearest friends who accompanied me through the long journey.
Last but certainly not least, I would like thank all the children and their parents for
their participation. My thanks also go to the funding source for this research: Alumni
Grants for Graduate Research and Scholarship from the Graduate School, Career
Development Grant from the Council of Graduate Students and Ruth Becky Irwin Fund
from the Department of Speech and Hearing Science at The Ohio State University.
Vita

2006  M.A. Experimental Phonetics, Graduate school, Chinese Academy of Social Sciences
2002  B.A. Chinese Language and Literature, Hunan University of Science and Technology

Publications


Field of study

Major Field: Speech and Hearing Science
# Table of Content

Abstract........................................................................................................................................... ii

Dedication.......................................................................................................................................... iii

Acknowledgments......................................................................................................................... v

Vita..................................................................................................................................................... vii

List of Tables.................................................................................................................................... xi

List of Figures................................................................................................................................... xiii

Chapter 1: General Introduction ................................................................................................. 1
References.......................................................................................................................................... 10

Chapter 2: Acoustic Development of Vowel Production in Native Mandarin-Speaking Children......................................................................................................................... 14
Abstract ........................................................................................................................................... 14
2.1 Introduction................................................................................................................................ 15
2.2 Methods....................................................................................................................................... 21
  2.2.1 Speakers .............................................................................................................................. 21
  2.2.2 Speech materials .............................................................................................................. 22
  2.2.3 Procedures ....................................................................................................................... 22
  2.2.4 Acoustic measurements ................................................................................................. 23
    2.2.4.1 Formant frequencies ............................................................................................. 23
    2.2.4.2 Vowel normalization ........................................................................................... 24
    2.2.4.3 Trajectory length ................................................................................................. 24
2.3 Results ......................................................................................................................................... 25
  2.3.1 Vowel duration ................................................................................................................ 25
  2.3.2 Midpoint F1 by F2 vowel space .................................................................................... 27
  2.3.3 Vowel space area ............................................................................................................ 31
  2.3.4 Formant dynamics .......................................................................................................... 33
    2.3.4.1 Formant movement pattern ............................................................................... 33
    2.3.4.2 Trajectory length ................................................................................................. 35
2.5 Discussion and Summary........................................................................................................... 41
References......................................................................................................................................... 46
Addendum.......................................................................................................................................... 53
List of Tables

Table 2.1. Summary of statistical results showing the effect of age on vowel duration of each Mandarin vowel (Means and standard errors are in ms). ......................................................... 26

Table 2.2. The area of ellipse of each Mandarin vowel in each age group (in Hz²)........... 29

Table 2.3. Summary of statistical results showing the effect of age on trajectory length of each Mandarin vowel (Means and standard errors are in Hz). .............................................. 37

Table 2.4. Skewness of the distribution curves of trajectory length for individual Mandarin vowels in each age group. ................................................................................ 40

Table 3.1. Summary of statistical results showing the effect of age on vowel duration of each English vowel (Means and standard errors are in ms).............................................. 65

Table 3.2. The ellipse areas of individual English vowels in each age group (in Hz²) .... 68

Table 3.3. Summary of statistical results showing the effect of age on trajectory length of each English vowel (Means and standard errors are in Hz).............................................. 74

Table 3.4. Skewness of the distribution curves of trajectory length for individual English vowels in each age group. ................................................................................ 80

Table 5.1. Characteristics of subgroups of bilingual Mandarin-English children: subject group, subject number, gender, age, length of residency in US, age of learning in English, English usage, age of learning in Mandarin, Mandarin usage. ............................................. 136

Table 5.2. Summary of statistical results showing the effect of L2 experience on vowel duration for each Mandarin vowel. (Means and standard errors of duration are in ms). 141

Table 5.3. The ellipse areas of individual Mandarin vowels in each group of children (in kHz²)......................................................................................................................... 144

Table 5.4. Euclidean distance between the centers of each vowel category of bilingual group with that of monolingual children for each Mandarin vowel. (Distances are in Hz) ......................................................................................... 145

Table 5.5. Summary of statistical results showing the effect of L2 experience on trajectory length of each Mandarin vowel (Means and standard errors are in Hz)........ 150
Table 5.6. Summary of statistical results showing the effect of L2 experience on vowel duration for each English vowel. (Means and standard errors are in ms) 152

Table 5.7. The area of ellipses of each English vowel in each group of children (in kHz²). 156

Table 5.8. The area of overlap for eight selected vowel pairs in each group of children (in kHz²). 156

Table 5.9. The proportion of each overlap to each vowel category for all eight vowel pairs in each group of children. 156

Table 5.10. Euclidean distance between the mean of each of bilingual group with that of monolingual children for each English vowel. 158

Table 5.11. Summary of statistical results showing the effect of L2 experience on trajectory length of each English vowel (Means and standard errors are in Hz). 164
List of Figures

Figure 2.1. Bar plots showing the mean and standard error of vowel duration for each Mandarin vowel in each age group. .............................................................. 26

Figure 2.2. Scatter plots of midpoint formant frequency values for individual Mandarin vowel categories in each age group. .......................................................... 28

Figure 2.2 continued ........................................................................................................ 29

Figure 2.3. Dispersion of Mandarin vowels for each age group plotted on the basis of the means and standard errors of rescaled normalized formant frequency values at midpoint location............................................................ 31

Figure 2.4. Box plots of Mandarin vowel space area for each age group. The vowel space areas of individual speakers within each group were also demonstrated. Each box depicts the lower quartile, median and upper quartile. Each asterisk represents a data point of one subject. ................................................................. 33

Figure 2.5. Vowel 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 five Mandarin vowels in the three age groups. The larger size symbol represents the 80% point. ....... 35

Figure 2.6. Bar plots showing the mean and standard error of trajectory length for each Mandarin vowel in each age group. ................................................................. 36

Figure 2.7. Probability density function of trajectory length for each Mandarin vowel of three age groups. ......................................................................................... 38

Figure 2.7 continued ..................................................................................................... 39

Figure 2.7 continued ..................................................................................................... 40

Figure 3.1. Bar plots showing the mean and standard error of vowel duration for each English vowel in each age group. ................................................................. 64

Figure 3.2. Scatter plot of midpoint formant frequency values for individual English vowel categories in each age group. ................................................................. 67
Figure 3.3. Dispersion of English vowels for each age group plotted on the basis of the means and standard errors of rescaled normalized formant frequency values at midpoint location................................................................. 69

Figure 3.4. Box plots of English vowel space area in each age group. The vowel space area of individual speaker within each group was also demonstrated. Each box depicts the lower quartile, median and upper quartile. Each asterisk represents a data point of one subject .............................................................................................................................. 70

Figure 3.5. Vowel dynamic 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 80% point .......... 72

Figure 3.6. Bar plots showing the means and standard errors of trajectory length for each English vowel in each age group ..................................................................................................................... 74

Figure 3.7. Probability density function of trajectory length for each Mandarin vowel across three age groups ........................................................................................................................ 76

Figure 3.7 continued ......................................................................................................... 77

Figure 3.7 continued ......................................................................................................... 78

Figure 3.7 continued ......................................................................................................... 79

Figure 1. Mean formant frequency values (with standard error) for L1 Mandarin (in black) and L2 English (in red) vowels produced by the child across eight epochs. Lines connect the three vowels /i, a, u/ in Mandarin (forming a triangular vowel space), the three vowels /ɪ, ð, ʌ/ in English in epoch 1 (a triangular vowel space similar to Mandarin) and the four corner vowels /i, æ, a, u/ in English from epoch 2 to epoch 8 (forming a quadrilateral space). ........................................................................................................ 101

Figure 4.2. Developmental trajectories for two back L2 English vowel pairs /ʊ, o/ and /a, ɔ/ produced by the child across eight epochs. Data points are redrawn from Figure 1. 105

Figure 4.3. Comparison of the child’s Mandarin and English vowel spaces at epoch 8 with the vowel spaces of corresponding monolingual adults. The left panel shows the dispersion of five monophthongal Mandarin vowels produced by monolingual Mandarin adult male speakers (data reported in Lin & Wang, 2001) and the dispersion of 11 nominal monophthongs in American English produced by monolingual English adult male speakers from the central Ohio area (data reported in Jacewicz, Fox & Salmons, 2007). In the middle panel, the child’s L2 English vowels are superimposed on those of monolingual English adults. In the right panel, the child’s L1 Mandarin vowels are superimposed on those of monolingual Mandarin adults. .............................................. 107
Figure 4.4. The child’s L1 Mandarin triangular vowel space (left) and L2 English quadrilateral vowel space (right) over eight epochs using rescaled normalized formant frequency values. .......................................................... 111

Figure 4.5. Regression model of the child’s rescaled normalized vowel space areas in L1 Mandarin (filled circles) and L2 English (unfilled circles). The unfilled triangle in epoch 1 is added to the plot to represent the early “English” vowel space defined by the three vowels /i, u, ʌ/. ................................................................................................................. 112

Figure 5.1. Bar plots showing mean and standard error of vowel duration for each Mandarin vowel in each group of children. ........................................................................................................... 141

Figure 5.2. Scatter plots of midpoint formant frequency values for individual Mandarin vowel categories in each group of children. ........................................................................................................ 143

Figure 5.3. Dispersion of Mandarin vowels for each group of children plotted on the basis of the means and standard errors of midpoint formant frequency values. ........................................ 145

Figure 5.4. Vowel dynamic 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 five Mandarin vowels in each group of children. The larger size symbols represent 80% point. ...................................................................................................... 146

Figure 5.5. Illustration of vowel dynamic spectral change for subsets of Mandarin vowels in each group of children. Note that the scale of each panel is different. The larger size symbols represent 80% point. ........................................................................................................ 148

Figure 5.6. Bar plots showing the mean and standard error of trajectory length for each Mandarin vowel in each group of children. .................................................................................................. 150

Figure 5.7. Bar plots showing mean and standard error of duration for each English vowel in each group of children ........................................................................................................ 152

Figure 5.8. Scatter plots of midpoint formant frequency values for individual English vowel categories in each group of children. .................................................................................................. 154

Figure 5.9. Illustration of overlap between /i/ and /e/ in Bi-low group of children…… 155

Figure 5.10. Dispersion of English vowels for each group of children plotted on the basis of the means and standard errors of rescaled normalized formant frequency values at midpoint location. ......................................................................................... 158

Figure 5.11. Vowel dynamic 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 in each group of children. The larger size symbols represent 80% point. ........................................................................................................ 159
Figure 5.12. Illustration of English vowel spectral change for four subsets of vowels in each group of children. Note that the scale of each panel is different. The larger size symbol represents 80% point.

Figure 5.13. Bar plots showing the mean and standard error of trajectory length for each English vowel in each group of children.

Figure 5.14. Mean of midpoint formant frequency values for Mandarin and English shared vowels in each group of children.

Figure 5.14 continued.

Figure 5.15. Vowel dynamic spectral change of Mandarin and English shared vowels for each group of children. Larger size symbol represent 80% point.
Chapter 1: General Introduction

During the first few years of life, most children can naturally acquire the magical tool separating human being from other animals--language. Although to most people, learning how to speak seems to be an inherited gift acquired without effort, the process of language acquisition is actually much more complicated than commonly thought. For the past few decades, our knowledge about the development of speech sound system from early infancy to the first few years of life has increased substantially. Researchers have outlined a developmental path from preverbal utterances (babblings) to meaningful speech streams. To date, the majority of previous literature concerning prespeech and speech development has provided much information about the perceptual development in infants as well as the relationship between perception and production in early childhood and the order of acquisition for individual speech sounds. However, relatively little research has focused on the development of speech after the early developmental period, especially the phonetic aspects of development in the children who have “acquired” the sound system but not yet reached the adult norms. As Bauman-Waengler (2000) pointed out, “…it appears that vowels are indeed generally mastered by the age of 3. Whether individual variation plays a large role in this acquisition process still needs to be documented. This is an interesting area of research, especially in light of the deviant vowel systems that are often noted in children with phonological disorders.” (pp.108) The
primary aim of this dissertation work is to further our understanding of the development of the acoustic properties of vowels in young children from diverse language backgrounds as well as the process of language separation and language interaction manifested in the phonetic systems in early sequential bilingual children.

In general, it has been claimed that vowel acquisition is basically completed by three years of age; significantly earlier than the mastery of consonants. Templin (1952, 1957) found that over 90% of vowels and diphthongs in words were produced correctly in children at three-year level, while Larkins (1983) reported that three-year children could produce all vowels with greater than 99% accuracy. However, many other studies found that vowel acquisition is not as straightforward and uniform as these studies might suggest. Actually, the acquisition of individual vowels follows different timelines. For example, Wellman, Case, Mengert and Bradbury (1931) found that two-year old children had only acquired /i, ɑ, u, o, ʌ, ə/, three-year old children mastered the additional vowels /ɛ, ɔ/ and the four-year old children acquired four more vowels /ɪ, e, æ, ʊ/. Davis and MacNeilage (1990) found that /u, o, i, ɪ/ were acquired earlier than /ɛ, ə/. Stoel-Gammon and Herrington (1990) provided the generalization that the corner vowels /i, ɑ, u/, the mid back /o/ and the central stressed /ʌ/ are the first group of vowels acquired, /æ, ʊ, ɔ, ə/ are the second group of vowels acquired. The front vowels /e, ɛ, ɪ/ and retroflexed vowel /ʃ, ʃ/ are the last group of vowels mastered.

These studies along with other literature on vowel acquisition agreed that certain vowels are acquired earlier than the other vowels. However, these studies diverged on the
specific order of acquisition of individual vowels as well as the specific accuracy rate of each vowel at particular age point. This divergence to a large extent is associated with the transcription-based methodology underlying the studies of early phonological development. On one hand, a phonetic transcription, especially narrow transcription derived from listeners’ subjective judgments, is affected by the listener’s training in phonetic transcription, their language experience, their basic skills and attention to detail, etc. On the other hand, some factors related to the speakers and the tasks such as children’s age, their physical and personality characteristics, the intelligibility of their speech and linguistic context may also affect the accuracy of transcription (Shriberg & Kent, 2003). Thus, phonetic transcription is not a direct, unbiased representation of the articulatory-phonetic properties of the speech sounds. To better understand the acoustical nature of the speech sounds in children rather than the perceptual quality on the basis of native listeners’ transcription, a more objective research method is needed.

The introduction of the spectrograph in 1946 and the progress of the acoustic analysis in the past few decades have enabled speech researchers to examine the acoustic characteristics of speech sounds in both adults and children. In 1952, Peterson and Barney conducted a seminal study in the acoustic features of vowel production in native English adults and children. This study provided a set of formant frequency values and fundamental frequency value in adult male, adult female and children speakers, which has been often cited as a referential base for later research. The authors reported that children have higher fundamental frequency and formant frequency values relative to adult male and female speakers. Since then, a large number of studies utilized this
technique to probe the phonetic aspects of infants’ vocalization and young children’s speech production.

Lieberman (1980) recorded five children’s utterance at 2-week intervals from 16 weeks to 5 years old and analyzed the acoustic features of the vocalizations (for young children) and vowels (for older children). He found that children’s formant frequencies were consistent with the vocal tract size. In particular, younger and smaller children had higher formant frequencies for specific vowels. When children grow and the length of their vocal tract increases, the formant frequency values fall. Gilbert, Robb & Chen (1997) examined the acoustic development of vocalization in infants aged between 15 to 36 months. They found that F1 range of the vocalizations was 910 -1321Hz at 15 months and decreased to 528 -900Hz at 36 months. The F2 range of the vocalizations was 2201-2733 Hz at 15 months and decreased to 1858- 2134 Hz at 36 months. Busby and Plant (1995) reported their data on acoustic features of 11 monophthongal Australian vowels at the steady-state produced by preadolescent boys and girls at four different ages (5, 7, 9, and 11 years old). The authors found that the formant frequencies of F1, F2, and F3 of all vowels decreased with increase in age. Additionally, the girls’ formant values were higher than those of boys.

These studies along with other acoustic-phonetic studies have converged on the finding of decreasing formant values as a function of chronological age (Assmann & Katz, 2000; Eguchi & Hirsh, 1969; Kent & Murray 1982; Lee, Potamianos & Narayanan, 1998; Perry, Ohde & Ashmead, 2001; Whiteside, Hodgson & Tapster, 2002). The decreasing pattern of formant frequencies resulted from the change of resonant features corresponding to the increasing size of vocal tract as a function of chronological age.
(Voperian & Kent, 2007). However, except for the developmental change of phonetic features as a consequence of increasing vocal tract, how are children’s phonetic features similar to or different from that of adults? Do they show continuing development in their acoustic features? Although previous studies claimed that children generally acquire the vowels of their native language by 3 years of age, due to the diverse “criteria” used to determine the acquisition of vowels in different studies, these “acquired” vowels are less likely to be fully developed and to achieve the adult norms. Some studies have suggested that the completion of a child’s speech sound system is not accomplished until 8 years of age (Kent, 1976; Sander, 1972; Tingley & Allen, 1975). But very little research has examined the continuing development of phonetic features of childhood speech production.

In addition, the majority of existing literature on childhood phonological and speech development has focused on English-speaking children, while relatively few studies had addressed speech development in children from other language backgrounds. As Zhu and Dodd (2006) stated “Multilingual studies contribute to our understanding of language acquisition by evaluating and challenging theoretical claims about development. Their value lies not only in their capacity to appraise claims about universal patterns of language acquisition, but also in their potential to examine whether and how differences in specific target languages result in differences in acquisition pattern.” (p. 3). In the book edited by Zhu and Dodd (2006), monolingual speech acquisition in a range of languages (including German, Putonghua, Cantonese, Maltese, Turkish, Spanish, Arabic and Mirpuri/Punjabi/ Urdu, etc) was examined. These studies followed the traditional transcription-based method in defining the sequential order of acquisition for individual
speech sounds and error patterns but provided no acoustic evidence regarding the
development of phonetic features in these children. To date, only a few studies conducted
cross-linguistic research to examine the language-specific acoustic characteristics in
childhood speakers. (Chung, Kong, Edwards, Weismre, Fourakis & Hwang, 2012; de
Boysson-Bardies, Halle, Sagart & Durand, 1989; Lee & Iverson, 2009; Rvachew,
Mattock, Polka, Menard, 2006).

A study by de Boysson-Bardies et al. (1989) compared the measures of mean F1,
mean F2, and F2/F1 ratio (which is an index of the compact-diffuse distinction between
vowels) of babble produced by 10-month-old infants from different language
backgrounds (Parisian French, London English, Hong Kong Cantonese and Algierian
Arabic). The comparable F2/F1 ratio between infants and the corresponding adults as
well as the different formant patterns among the infants from diverse languages showed
that infants as young as 10 months of age already demonstrate language-specific acoustic
patterns. More recently, Rvachew et al. (2006) reported their data on the formant patterns
in 10- and 18-month-old infants exposed to Canadian English (CE) or Canadian French
(CF). The authors found that infants exposed to CE showed a decreasing mean F2 while
infants exposed to CF showed a stable mean F2. In addition, although both CE and CF
infants showed decreasing mean F1, the magnitude of decrease in CF infants was greater
than that in CE infants. Chung et al. (2012) examined the relative position of three shared
corner vowels /a, i, u/ across five languages (American English, Cantonese, Greek,
Japanese and Korean) in 2-year, 5-year-old children, and adults. They found that both
children and adults showed language-specific features in terms of the location of these
shared corners vowels and the shape of acoustic space connected by these three corner vowels.

These studies provided evidence regarding the development of language-specific phonetic features from infants to young children and the influence of language exposure on the early stages of speech development. For the most part, these studies have examined monolingual children. How do individual phonetic features develop in bilingual children who are exposed to more than one language during the process of their language acquisition; how do they deal with the language-specific features in different languages; how do they organize the different language systems; and how do the two language systems interact with each other? Answering these questions can not only further our understanding of speech development in young bilingual children, but also shed new light on the mechanism of children’s language acquisition.

To date, there are a substantial number of studies which have examined the acoustic features of adult bilingual speakers who differ in the age of learning, amount of L1/L2 usage, etc. (Aoyama, Guion, Flege, Yamada, & Akahane-Yamada, 2008; Baker, Trofimovich, Flege, Mack, & Halter, 2008; Bohn & Flege, 1992; Flege, Frieda & Nazawa, 1997; Guion, 2003; McAllister, Flege, & Piske, 2002; Munro, Flege, & MacKay 1996). Although examining adulthood bilingual speakers’ phonetic features enables us to understand the ultimate language attainment as a consequence of interaction of multiple factors, we still lack information regarding the ongoing process of speech development in early bilingual children or direct comparison of bilingual children’s speech development with that of monolingual peers.
For example, Dodd, So and Li (1996) suggested that bilingual children may go through a qualitatively different development path in phonological acquisition than monolingual children do. In addition, regarding the long time debate about whether there are one or two systems in bilingual children, some researchers believe that while simultaneous bilingual children go through a unitary system stage (Oksaar, 1971; Schnitzer & Krasinski, 1994; Volterra & Taeschner, 1978), sequential bilinguals may use one phonological system as the foundation and separate the second language system from the first one by modifying or adding elements to the first language system. Moreover, in terms of another widely researched topic on the interaction effect between L1 and L2 in bilingual speakers, some researchers claimed a more active interference and bi-direction interaction in children relative to adults (Baker & Trofimovich, 2005; Flege, Schirru & MacKay, 2003). However, the direction of the L1 and L2 sound changes due to such interaction effects remains controversial. To address these questions, a closer investigation of bilingual children rather than fully developed adult bilinguals is certainly needed.

The objective of this dissertation is two-fold: The first aim is to investigate the development of vowel production acoustically in monolingual Mandarin1 and English children relative to monolingual adults. In particular, we plan to unveil the continuing development of acoustic properties in relatively young children to fill the gap of knowledge in acoustic studies in early childhood. The second aim is to document the separation of L1 and L2 phonetic systems and examine the effect of L1-L2 interaction on the phonetic features of bilingual Mandarin-English children relative to those of monolingual Mandarin-English children.

---

1 In this dissertation, we use the term ‘Mandarin’ to refer to Standard Chinese (Putonghua).
monolingual children. To accomplish these goals we conducted both 1) a longitudinal study to provide a complete development path of phonetic details in one emergent bilingual child and 2) a cross-sectional study to examine the effect of L1-L2 interaction on early bilingual children’s acoustic manifestations.

The organization of the dissertation is as follows: Chapter 2 examines the development of static and dynamic vowel features in monolingual Mandarin children aged 3-6 years and compare the children’s acoustic parameters with those of monolingual Mandarin adults. Chapter 3 examines the development of both static and dynamic vowel features in monolingual English children aged 3-6 years and compared their acoustic parameters with those of monolingual English adults. Chapter 4 reports a 20-month longitudinal case study on the acoustic development of vowels in an emergent bilingual Mandarin-English speaking child. Chapter 5 investigates the static and dynamic vowel features in 5-6 year-old bilingual Mandarin-English children compared to age-matched monolingual children selected from Chapters 2 and 3. Chapter 6 provides a general discussion regarding the speech development of early monolingual and bilingual children.
References


Chapter 2: Acoustic Development of Vowel Production in Native Mandarin-Speaking Children

Abstract

The present study aims to document the developmental profile of both static and dynamic acoustic features of vowel productions in monolingual Mandarin-speaking children between 3 and 6 years of age. 29 monolingual Mandarin children and 12 native Mandarin adults were recorded producing ten Mandarin disyllabic words containing five basic monophthongal vowel phonemes /a, i, u, y, ɤ/. The frequency values of F1 and F2 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 but the size of individual vowel categories exhibited a decreasing trend as a function of increased age. 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. Each Mandarin vowel was characterized by a distinctive formant movement pattern. Mandarin children generally demonstrated formant movement patterns comparable to those of adult speakers. However, children still showed positional variation for certain vowels. They also differed from adults in the magnitudes of spectral change for certain vowels. This indicates the vowel development is a protracted process which extends beyond 3 years of age.
2.1 Introduction

As opposed to the large number of studies on consonants, relatively few studies focused on children’s vowel acquisition and development, especially for non-English speaking children. This is probably due to the claims in previous research that vowels are relatively easier to acquire and are acquired at an early stage of childhood, usually earlier than consonants (Templin, 1957; Smith, 1973; Waterson, 1971; Zhu & Dodd, 2000). However, vowel acquisition is not as straightforward as it has been thought. On one hand, the acquisition of individual vowels follows different paths. Some vowels are acquired earlier, while the other vowels are acquired relatively late (Davis & MacNeilage, 1990; Otomo & Stoel-Gammon, 1992; Wellman, Case, Mengert, & Bradbury, 1931). On the other hand, even typically-developing children exhibit great individual differences in phonological development (Vihman, 1993; Vihman, Gerguson & Elbert, 1986). Not every child follows the same path in acquiring phonetic inventories in his/her native language. The process of vowel development deserves further examination. Furthermore, within the limited number of studies on children’s vowel development, a majority of them have targeted English-speaking children, resulting in a relatively paucity of research in children’s vowel development from other language backgrounds. The present research focuses on the development of phonetic features of vowel production in native Mandarin-speaking children. In particular, we will examine the acoustic features including vowel duration, formant frequencies, vowel space area and dynamic vowel spectral change to document the acoustic profile of vowel development in Mandarin-speaking children.

Traditionally, research on children’s vowel acquisition has focused on the order of acquisition of individual vowels on the basis of accuracy rate of children’s speech
production transcribed and evaluated by native adult speakers (Davis & MacNeilage, 1990; Hare, 1983; Larkins, 1983; Otomo & Stoel-Gammon, 1992; Paschall, 1983; Templin, 1952, 1957; Wellman et al., 1931). Indeed, the transcription-based accuracy rating method provides direct evaluation of children’s phonological skills. In addition, based on the perceived quality of children’s vowel production, adult listeners can determine whether the vowels are distinguishable enough to be classified as different vowel categories and thus generalize the acquisition order of vowel phonemes. However, transcribing children’s speech samples is affected by a number of factors including children’s age, children’s physical and personality characteristics, intelligibility of children’s speech, linguistic context, etc. (Shriberg & Kent, 2003). Furthermore, the adult transcribers do not always agree with each other in their transcriptions. In addition, the transcriptions often do not provide the fine phonetic details of children’s speech production. Therefore, a more objective research approach is needed to better examine the acoustic characteristics of children’s speech development.

A substantial body of previous research has investigated the developmental change of acoustic features in native English-speaking children across different age spans (Buhr, 1980; Busby & Plant, 1995; Eguchi & Hirsh, 1969; Gilbert, Robb & Chen, 1997; Lee, Potamianos & Narayanan, 1999; Lieberman, 1980; Kent & Murray, 1982; Kuhl & Meltzoff, 1996). Children’s vocal tract grows as a function of their chronological age and 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, 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. However, it remains unclear if the effects produced by physical changes to 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. In this case, if a child’s acoustic patterns are the same as those of adults, it indicates that the child can produce the vowel sounds in an adult-like manner. If a child’s acoustic features show different patterns from those of adults, 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 developed the vowels yet.

To date, most studies examining the acoustic properties of children’s vowel production have focused on the midpoint value of formants and its position in the F1 x F2 plane to characterize the acoustic nature of the vowel. The midpoint presumably represents the steady state portion of the vowel. However, previous literature shows ample evidence that vowels are more precisely characterized by inherent spectral change (Harrington & Cassidy, 1994; Hillenbrand, Getty, Clark and Wheeler, 1995; Neary & Assmann, 1986). For example, Hillenbrand et al. (1995) found that nearly all English vowels showed some measurable degree of formant movement regardless of their consonant neighbors. Vowel spectral change also conveys important information associated with speaker’s dialect (Fox & Jacewicz, 2009) and speaker’s generation (Jacewicz, Fox & Salmons, 2011a). Furthermore, vowel spectral change is also related to the development of children’s speech motor control. Both temporal and spectral measures have long been used as indices to estimate speech motor development in children
(DiSimoni, 1974; Kent & Forner, 1980; Nittouer, 1993; Smith 1991). Specifically, during the process of vowel production, the change of the vocal tract shape resulted from the movement of articulators may cause measurable change of formant frequency values. Therefore, due to the less mature ability to control articulators, children may produce different formant frequency patterns compared to adult norms even after the factor of maturation has been removed by vowel normalization. The dynamic vowel spectral change associated with the ongoing coordination of articulators may also demonstrate different patterns from those of adults as a consequence of the development of children’s speech motor control. The apparent importance of vowel spectral change motivates us to further explore the dynamic features of childhood vowel production to better understand the nature of speech sounds in children. However, only a few studies have examined the development of vowel dynamic spectral change.

Jacewicz, Fox and Salmons (2011b) investigated vowel spectral change in older English children aged 8 to 12 years across northern, midland and southern dialects of American English. They found that children at this age span have comparable patterns of vowel spectral change as adults in their own dialect. More recently, Assmann, Nearey and Bharadwaj (2013) examined the pattern of vowel spectral change in children aged 5 to 18 years from the southern American English region. They also found that children as young as five years old already exhibited consistent vowel spectral change patterns similar to those of adults. These studies provided basic knowledge regarding the acoustic features of vowel spectral change in native English-speaking children older than 5 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 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.

The present study aims to expand the current knowledge of children’s vowel development by systematically examining both static and dynamic features of vowel production in native Mandarin-speaking children. Mandarin is the official language spoken in China and used by approximately 1.5 billion speakers all over the world. To our knowledge, there is relatively little published research on Mandarin-speaking children’s vowel development. 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; Si, 2006; Zhu & Dodd, 2000). Recently, Shi and Wen (2007) examined the acquisition of 7 Mandarin vowels /a, i, u, y, ə, ɿ, ɿ/ among 40 native Mandarin-speaking children aged between 1 to 6 years. In addition to providing the acquisition order of individual vowels on the basis of evaluating the accuracy rate of the children’s vowel production, this study provided several samples of individual children’s acoustic vowel space to characterize the nature of their speech errors. 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 age ranges.
On the other hand, regarding the acoustic features of Mandarin vowels in native Mandarin-speaking adults, previous studies have examined the basic features such as vowel duration and formant frequencies at the static portion (represented at the midpoint) for monophthongs and patterns of formant movement for diphthongs and triphthongs (Cao & Yang, 1984; Howie, 1976; Wu, 1986; Zee, 2001). In particular, Howie (1976) measured formant frequencies at multiple time locations over the duration of the vowel. The number of measurement point varied as a function of the size of the expected formant movement. For the single vowels such as /a/ and /u/, only midpoint (50% point) formant frequencies were measured. For the single vowels which show gradual formant movement such as /i/ and /y/, he measured formant frequencies at the vowel onset and midpoint. For the compound vowels, he measured formant frequencies at three, four or five equally-spaced points over the vowel duration. Howie attempted to investigate the spectral change in Mandarin vowels. However, since he selected different measure locations on the basis of the extent of observed formant movement on the spectrogram, he did not systematically examine the nature of spectral change across all vowels especially all single vowels. The present study, in many ways, is an extension of this seminal work about Mandarin vowel dynamics. In addition, the present study will also apply the acoustic measurements widely used in English vowel studies (Fox & Jacewicz, 2009; Jin & Liu, 2013) to quantitatively define the nature of spectral change in Mandarin vowels.
The present study has three related research aims:

1. To document the developmental profile of static and dynamic acoustic features of five basic vowels in Mandarin speaking children aged between 3 to 6 years of age;
2. To examine to what extent the acoustic characteristics of children’s vowels are similar to or different from those of Mandarin-speaking adults; and
3. To provide a detailed profile of vowel spectral change in native Mandarin-speaking adults.

2.2 Methods

2.2.1 Speakers

The speakers included 29 native Mandarin-speaking children (13 females and 16 males) aged 3 to 6 years and 12 native Mandarin-speaking adults (6 females and 6 males) aged 23 to 58 years. These speakers fell into three age groups: younger children (AY: 3-4, 14 children), older children (AO: 5-6, 15 children) and adults (AA). All child speakers were born in Mandarin-speaking regions and raised in the Beijing area, where the data collection occurred. Both parents of all children speak Mandarin in daily life. All Mandarin-speaking adults were born and raised in Mandarin-speaking regions. They were also recruited from the Beijing area. Five of twelve adults were the teachers employed in the kindergarten where the child speakers were recruited. All speakers were reported as having no speech or language disorders.
2.2.2 Speech materials

The speech materials (see Addendum) included ten Mandarin disyllabic words containing five Mandarin monophthongal vowel phonemes /a, i, u, y, ɤ/. Each vowel occurred in two different words. In each disyllabic word, the target vowel always appeared in the first syllable. The preceding consonants of the target vowels were stop consonants except for /y/. The target syllable had no consonant in its coda. The initial consonant in the second syllable was not strictly controlled. In addition, as shown in previous studies (Chang, 2010; Lin, 1988; Tseng, 1981, Xu, Tsai & Pfingst, 2002), vowels in a syllable with tone three have a significantly longer duration than those produced in a syllable with one of other three tones. Thus, no target syllable contains tone three. All words were selected based on their picturability and familiarity to young children.

2.2.3 Procedures

The word productions were recorded through a visual-auditory word repetition task in a quiet room under the control of a custom MATLAB program. During the recording period, each speaker was seated in front of a laptop computer wearing a Shure SM10A head-mounted microphone situated approximately 1-inch from the speaker’s

---

2 Lin and Wang (2001) identified eight vowels (/i, ɨ, ɿ, y, u, a, ɤ, o/) in the Mandarin vowel system. Among these eight vowels, ɨ/ and ɿ/ are allophonic variations of the vowel /i/. /o/ is often considered a diphthong rather than a monophthong (Zee, 2001). Therefore, these three vowels are not included in the present study.

3 Due to the phonetic rules in Mandarin syllable structure, /y/ cannot be preceded by a stop. Therefore, the syllables containing /y/ has either zero initial (no preceding consonant) or the affricate /ʨ/.
mouth. For each speaker, the recording session was divided into two blocks. In each block, pictures corresponding to target words were randomly ordered⁴ and presented on the computer screen. An audio prompt⁵ for each word produced by a native Mandarin speaker was played to each speaker. The speaker was then asked to repeat the word once immediately after the audio prompts. Speech samples were recorded directly onto a hard disk of the laptop with a 16-bit quantization rate and 44.1 kHz sampling rate.

2.2.4 Acoustic measurements

2.2.4.1 Formant frequencies

All tokens were down-sampled to 11.25 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, Fox & Salmons, 2011c; Jacewicz, Salmons & Fox, 2009), formant frequencies are measured at five equidistant temporal locations (20-35-50-65-80% point) to capture the dynamic spectral change of the vowel movement. The first and fifth point for the vowel acoustic measurements were selected at 20% and 80% over the vowel duration to avoid the influence of preceding and following consonants. The temporal locations were determined based on 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

---

⁴ The same random order was used for both recording blocks.
⁵ The audio prompts were produced in a citation form. For the words with a neutral tone in the second syllable, they were produced in a manner to avoid the shortening in the second syllable as they are in spontaneous speech.
period of voicing following stop closure release or cessation of frication. For the vowel /y/ in “yú” (fish), the vowel onset 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.

2.2.4.2 Vowel normalization

To eliminate the effect of vocal tract length among speakers, normalized formant frequency values were calculated using the method in Lobanov (1971) (found to be one of the most effective normalization methods, Adank, Smits & van Hout, 2004). Since the normalized formant frequency values represent z scores, they were then rescaled into Hz-like values using the method proposed by Thomas and Kendall (2007) to facilitate interpretation of normalized vowel spaces. In particular, $F_1' = 250 + 500 \left( \frac{F_{N1} - F_{N1\text{MIN}}}{F_{N1\text{MAX}} - F_{N1\text{MIN}}} \right)$ and $F_2' = 850 + 1400 \left( \frac{F_{N2} - F_{N2\text{MIN}}}{F_{N2\text{MAX}} - F_{N2\text{MIN}}} \right)$ where $F_i'$ is a rescaled, normalized formant; $F_{N_i}$ is a Lobanov normalized formant value; and $F_{N_i\text{MIN}}$ and $F_{N_i\text{MAX}}$ are the minimum and maximum values of Lobanov normalized $F_{N_i}$ across the entire dataset.

2.2.4.3 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). It is 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%),

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

(1)
where the length of each vowel section (VSL) is calculated based on the formula:

\[ VSL_n = \sqrt{(F_{1n+1} - F_{1n})^2 + (F_{2n+1} - F_{2n})^2} \]  \hspace{1cm} (2)

The assumption is that a longer trajectory length indicates a greater change of formant frequency values and greater magnitude of formant movement. This measure provides a detailed assessment of the magnitude of formant movement over the course of vowel duration between the selected time locations particularly for the curved formant track which has formant frequency values in later temporal portions bending over to those in the former temporal portions.

2.3 Results

2.3.1 Vowel duration

Vowel duration was determined from the manually located onset and offset landmarks. As shown in Figure 2.1 and Table 2.1, vowels produced by children are generally longer than those produced by adults, consistent with the findings in previous studies (e.g., Lee et al. 1999). The effect of vowel quality on duration was reported in previous literature (Howie, 1976), one-way ANOVA was used to test the effect of age on each vowel. Then, Tukey HSD was used to test the nature of significant differences between each two age groups. The results showed that adults produced significantly shorter vowel duration for the vowel /i/ and /ɤ/ than both younger and older children did.

For the vowel /u/, adults showed significant shorter vowel duration than older children.
Figure 2.1. Bar plots showing the mean and standard error of vowel duration for each Mandarin vowel in each age group.

<table>
<thead>
<tr>
<th></th>
<th>AY M (SE)</th>
<th>AO M (SE)</th>
<th>AA M (SE)</th>
<th>p</th>
<th>$\eta^2$</th>
<th>Significant difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>296 (16)</td>
<td>275 (12)</td>
<td>249 (17)</td>
<td>0.117</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>296 (12)</td>
<td>299 (10)</td>
<td>252 (16)</td>
<td>0.004</td>
<td>0.172</td>
<td>AA &lt;AY, AA &lt;AO</td>
</tr>
<tr>
<td>u</td>
<td>258 (11)</td>
<td>271 (12)</td>
<td>227 (14)</td>
<td>0.048</td>
<td>0.147</td>
<td>AA &lt;AO</td>
</tr>
<tr>
<td>y</td>
<td>303 (17)</td>
<td>293 (13)</td>
<td>260 (14)</td>
<td>0.157</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>317 (13)</td>
<td>330 (11)</td>
<td>267 (14)</td>
<td>0.004</td>
<td>0.253</td>
<td>AA &lt;AY, AA &lt;AO</td>
</tr>
</tbody>
</table>

Table 2.1. Summary of statistical results showing the effect of age on vowel duration of each Mandarin vowel (means and standard errors are in ms).
2.3.2 Midpoint F1 by F2 vowel space

Traditionally, midpoint formant frequency values have been used to characterize the basic acoustic feature of vowels. In the present study, the rescaled normalized F1 and F2 at midpoint location were used to plot the vowel dispersion pattern (shown in Figure 2.2). Vowel ellipses were plotted in a manner similar to that used in Zhou and 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 the semimajor axis at the center was defined as the semiminor axis. The lengths of the two axes were determined by two standard deviations of all data points away from the center along the respective lines. In this case, each ellipse encircled approximately 95% of the data points for each vowel category.

As shown in the plot for adults, there are five clearly separated vowel phoneme categories in each of which, the speech samples were highly concentrated. The vowels /a, i, u/ were located in the corner positions. /y/ was close to /i/ while /ʌ/ was located in a relatively central position. The definable vowel categories were also clearly presented in both groups of children, which indicated that the children in the present study could also produce separate vowel categories for these five vowel phonemes. Although all child speakers seemed to differentiate 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 shown in Table 2.2. The sizes of ellipses decreased as a function of speakers’ age for all five vowel categories. Therefore, although the child speakers can separate these five vowel categories in the acoustic vowel space, they still demonstrated greater variation in the formant frequencies than did adults but the amount of variation decreased as the child speakers grew older.

Figure 2.2. Scatter plots of midpoint formant frequency values for individual Mandarin vowel categories in each age group.

Continued
Table 2.2. The area of ellipse of each Mandarin vowel in each age group (in Hz^2).

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>y</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY</td>
<td>19,097</td>
<td>18,581</td>
<td>20,376</td>
<td>28,841</td>
<td>34,642</td>
</tr>
<tr>
<td>AO</td>
<td>18,062</td>
<td>8,777</td>
<td>8,968</td>
<td>17,408</td>
<td>17,591</td>
</tr>
<tr>
<td>AA</td>
<td>9,861</td>
<td>4,122</td>
<td>5,699</td>
<td>11,373</td>
<td>8,733</td>
</tr>
</tbody>
</table>

Figure 2.3 shows the means and standard deviations of 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 all Mandarin vowels and the deviation resided in the F2 dimension for most vowels. In order to better examine to what extent the relative position of each vowel changed as a function of chronological age, One-way ANOVA was used separately on rescaled normalized F1 and F2 across three age groups for each vowel. The results showed that for the vowel /a/,
there was a significant age difference on F2 (F(2,38) = 4.339, p = 0.02, \( \eta_p^2 = 0.186 \)). In particular, older children produced higher F2 than did adults. For the vowel /i/, there was a significant age difference on F2 (F(2,38) = 10.54, \( p < 0.001, \eta_p^2 = 0.357 \)). In particular, both younger and older children produced higher F2 than did adults. For the vowel /u/, there was a significant age difference on F1 (F(2,38) = 5.764, \( p = 0.007, \eta_p^2 = 0.233 \)). Specifically, older children produced lower F1 than did adults. For the vowel /y/, the statistical result revealed a significant age effect on F2 (F(2,38) = 8.173, \( p = 0.001, \eta_p^2 = 0.301 \)). In particular, both younger and old children produced lower F2 than did adults. As for the vowel /\u0103/, the statistical result showed significant age difference on F2 (F(2,38) = 3.694, \( p = 0.034, \eta_p^2 = 0.163 \)). In particular, older children produced lower F2 than did adults. These statistical results indicate that children at this age range still differed from adults in the static features of Mandarin vowels even though the influence of different vocal tract size had been removed.
Figure 2.3. Dispersion of Mandarin vowels for each age group plotted on the basis of the means and standard errors of rescaled normalized formant frequency values at midpoint location.

### 2.3.3 Vowel space area

The working vowel space is defined as the area surrounded by the boundary vowels which are assumed to represent the most peripheral positions of the vowel space and correspond to the maximum articulatory positions that can be reached by a speaker. The overall size of vowel working space is an important index to illustrate the developmental change of children’s vowel production as a function of increase in vocal tract length (Voperian & Kent, 2007). It is also used to examine the nature of vowel structure differences across different languages (Chung et al., 2012) as well as different dialects of speakers (Fox & Jacewicz, 2010). In addition, vowel working space is closely associated with speech intelligibility. Recent studies showed that speakers with
communication disorders caused by dysarthria, neuromuscular disease such as amyotrophic lateral sclerosis and some neurogenic diseases such as cerebral palsy produced significantly reduced vowel working space area compared to normal speakers (Higgins & Hodge, 2001; Liu, Cao & Kuhl, 2005; Weismer, Jeng, Laures, Kent & Kent, 2001). These studies, collectively, indicate that vowel working space area may represent an important characteristic of vowel production.

The most common approach taken 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 represent 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 2.4. The one-way ANOVA result revealed no significant difference among these three groups of speakers in terms of the normalized vowel space area (F(2,38)=1.392, \( p > 0.05 \), \( \eta^2_p = 0.068 \)). This result indicates that Mandarin children at this age range had developed an adult-like basic vowel framework.
2.3.4 Formant dynamics

2.3.4.1 Formant movement pattern

The spectral dynamics patterns for these three age groups are shown in Figure 2.5. For adult speakers, the five Mandarin vowels demonstrated distinctive formant trajectories. In particular, for the two high front vowels /i/ and /y/, the vowel /i/ remained relatively stable in both F1 and F2, which 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 50% to 80% point. The other vowel /ɤ/ demonstrated substantial increase in F1 but relatively no change in F2. In terms of the direction of formant trajectories, the offset of all five vowels pointed toward the central position of the vowel space.

The child speakers showed comparable formant movement patterns to the adults for most vowels. However, for the vowel /i/, the format trajectory produced by children showed greater formant movement than those produced by adults. In addition, the vowel /i/ produced by child speakers was located at a more fronted position than that produced by adult speakers. The vowel /y/ demonstrated a gradual forward movement as a function of age. As shown in the vowel space, the proximity of /i/ and /y/ in the production of the adults is greater than that of the children. For the other two vowels /ɤ/ and /u/, the vowel /ɤ/ was fronted while the vowel /u/ was slightly lowered from children to adults.
Figure 2.5. Vowel 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 five Mandarin vowels in the three age groups. The larger size symbol represents the 80% point.

2.3.4.2 Trajectory length

Other than the relative position of individual vowels across these three age groups, the magnitude of formant movement also differs. 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 2.6). Table 2.3 provides a summary of these values for each vowel in each age group. For adult speakers, among these five Mandarin vowels, /u/ had the longest trajectory length while /i/ had the shortest trajectory length. In general, adult speakers have shorter trajectory lengths than child speakers for the vowels /a, i, y, ɤ/ but
not for /u/. The effect of vowel quality on trajectory length is anticipated and is not of particular interest in the present study. One-way ANOVA was used to test how the age factor influences the trajectory length for each vowel. Tukey HSD was used to examine pairwise differences between each two age groups. The results show that for the vowel /y/ and /ɤ/, adults have significantly shorter trajectory lengths than older children while for the vowel /i/, adults have significantly shorter trajectory lengths than both younger and older children.

Figure 2.6. Bar plots showing the mean and standard error of trajectory length for each Mandarin vowel in each age group.
Table 2.3. Summary of statistical results showing the effect of age on trajectory length of each Mandarin vowel (means and standard errors are in Hz).

<table>
<thead>
<tr>
<th></th>
<th>AY M (SE)</th>
<th>AO M (SE)</th>
<th>AA M (SE)</th>
<th>p</th>
<th>η²</th>
<th>Significant difference</th>
<th>Tukey HSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>178.70 (11.71)</td>
<td>148.50 (12.44)</td>
<td>139.75 (9.98)</td>
<td>0.061</td>
<td>0.137</td>
<td>AA &lt;AY, AA &lt; AO</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>161.40 (18.33)</td>
<td>184.48 (11.20)</td>
<td>104.17 (15.22)</td>
<td>0.002</td>
<td>0.274</td>
<td>AA &lt;AY, AA &lt; AO</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>191.30 (11.01)</td>
<td>176.45 (9.39)</td>
<td>195.05 (12.96)</td>
<td>0.443</td>
<td>0.042</td>
<td>AA &lt; AO</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>135.86 (9.90)</td>
<td>145.37 (8.08)</td>
<td>115.11 (7.65)</td>
<td>0.05</td>
<td>0.146</td>
<td>AA &lt; AO</td>
<td></td>
</tr>
<tr>
<td>ɤ</td>
<td>154.77 (6.82)</td>
<td>159.02 (6.28)</td>
<td>134.42 (4.71)</td>
<td>0.02</td>
<td>0.185</td>
<td>AA &lt; AO</td>
<td></td>
</tr>
</tbody>
</table>

To further examine whether the trajectory lengths in children were similar to or different from adults’ norm, the distribution curves of trajectory length of three age groups for each Mandarin vowel were generated using the kernel density estimation function in Matlab and shown in Figure 2.7. In general, kernel density estimation generates a smooth probability density function on the basis of a finite data sample to refer to the distribution of the unknown target population. This method is closely related to a histogram but overcomes the limitation of discontinuity and dependence on the end point of each bin associated with the histogram by introducing two 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. Therefore, it 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 this figure, the trajectory lengths of the vowels /i, y/ in children were distributed with greater variability than those in adults. In addition, the
distribution curves of these two vowels in children also displayed slight mismatch on peak location from those in adults. These deviations of children’s distribution curve from adults’ target suggest that children at this age had not developed an adult-like manner in terms of the magnitude of formant movement. The distribution curves of the vowel /ɤ/ and /a/ showed gradual change from younger children to adults, which indicates the approximation of children’s trajectory lengths to adults’ target. However, for the vowel /u/, younger children group showed highly similar distribution curve as adults. The curve of older children even showed less variability than that of adults. This indicates that the children at this age had developed an adult-like spectral change feature in terms of the magnitude of the formant movement.

Figure 2.7. Probability density function of trajectory length for each Mandarin vowel of three age groups.
Figure 2.7 continued

Continued
It should be noted that not every group of speakers had a normal distribution curve for every vowel. As shown in Table 2.4, 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 some vowels such as /a/ produced by younger children, /i/ produced by older children and adults demonstrated bimodal distributions. This information is not evident from the means and standard errors. Thus the density function analyses provide important evidence regarding the developmental change of the trajectory length which may not be appropriately shown by the ANOVA test.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>y</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY</td>
<td>0.850</td>
<td>0.973</td>
<td>0.537</td>
<td>0.368</td>
<td>0.461</td>
</tr>
<tr>
<td>AO</td>
<td>0.234</td>
<td>0.871</td>
<td>0.413</td>
<td>1.131</td>
<td>0.322</td>
</tr>
<tr>
<td>AA</td>
<td>0.217</td>
<td>1.737</td>
<td>0.560</td>
<td>0.150</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Table 2.4. Skewness of the distribution curves of trajectory length for individual Mandarin vowels in each age group.
2.5 Discussion and Summary

The main purpose of this study was to examine the development of the acoustic properties of vowel production in Mandarin-speaking children and how these acoustic features are similar to or different from adults’ norms after eliminating the effects of vocal tract length. A series 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. 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, which indicated that child speakers as young as 3 years can separate these five vowel categories well in the acoustic vowel space. However, the size of the vowel clusters exhibited a decreasing trend as a function of increased age. In addition, children also showed statistical differences from adults in the rescaled formant frequency values at midpoint for all five Mandarin vowels. These results suggested that the children in this age range are still showing developmental change in terms of the fine acoustic features of their vowel production. The normalized vowel space area demonstrated similar patterns between children and adults, which indicates that children at this age range are able to establish the basic vowel acoustic frame as do adults.

To better understand the nature of acoustic features of vowel production in Mandarin children, we also examined their dynamic vowel spectral change, which was compared with the adults’ targets. The spectral change of the five Mandarin vowels for adults demonstrated that each Mandarin vowel is characterized by a distinctive formant movement pattern in terms of the shape and movement direction of the formant.
trajectories. In particular, /a/ showed a bowed formant trajectory due to the reversal change in F1. /i/ demonstrated a relatively static formant trajectory. /u/, /y/ and /u/ displayed relatively straight formant trajectories but the direction of the formant trajectories are distinctive. For all five vowels, the offset of their formant trajectories points to the central region of the vowel space. This may result from the syllable structure in Mandarin. In general, except for the syllables that ended with the nasal sounds /n/ or /ŋ/, all other Mandarin syllables are ended with vowels. Basically, the vowel located at the syllable ends tends to move toward the central position (like a schwa) at the offset of the vowel duration. This results in the centralized tendency of the vowels displayed in the acoustic vowel space. Compared to adults’ vowel formant trajectories, Mandarin-speaking children generally demonstrate a comparable formant movement pattern. However, certain vowels showed evident positional change from children to adults. This suggests the continuing development of these vowels from children to adults.

The establishment of the profile of vowel spectral change in Mandarin can be of more interest to future studies on accented Mandarin. It is known that Mandarin is the basic dialect of Standard Chinese, the official language used in China. Most people from other dialect regions speak Mandarin with an accent. With the patterns of dynamic vowel spectral change specified in native Mandarin speakers, more detailed comparison can be derived to help us comprehensively understand in which way the accent Mandarin is similar to or different from native Mandarin as well as how the Mandarin vowels produced by people from different dialects will be affected by their native dialects.
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. For adult speakers, /u/ has the longest trajectory length while /i/ has the shortest trajectory length. This indicates that /u/ has greater magnitude of formant movement while /i/ is relatively stationary. Regarding the formant dynamic measurements in children, our child speakers generally showed greater trajectory length and relatively larger variation than adults. This finding suggested that adults’ productions of these monophthongal vowels were more stable than those of children. The longer trajectory length in child speakers may partially be related to their longer vowel duration. When speakers produce a vowel with longer duration, the articulators may show more movement which will cause greater formant fluctuation (Jacewicz, Salmons & Fox, 2009).

In addition to the longer vowel duration, the longer trajectory length in children may also result from their relatively less control of the articulators compared to adults. Previous literature has shown that children move their articulators slower (Smith & Gartenberg, 1984) and exhibit much greater variability in the movement of articulators than do adults (Smith, 2006). In addition, it has also been 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 (Green, Moore & Reilly, 2002; Sharkey & Folkins, 1985; Smith, 2006; Smith, Gartenberg, 1984; Smith & Goffman, 1998; Smith & Zelaznik, 2004; Walsh & Smith, 2002). Specifically, Smith & Zelaznik (2004) found that children showed increasing
improvement in their speech motor skills, which reached a plateau from 7 to 12 years of age with continuing development at the adolescence. Green et al. (2002) found that the development of motor control of different articulators is in a sequential manner. In particular, the articulatory control of jaw develops earlier than that of lips in young age children. According to these findings, it is very likely that due to the continuing development of speech motor skills, children aged 3 to 6 may not be able to keep the articulators as stable as adults when producing monophthongal vowels. This caused the greater formant movement in their vowel production. The different pattern of density function of trajectory length between children and adults in certain vowels also evidenced the process of development of children’s trajectory length to adult norms. However, not every vowel showed developmental change from children to adults in terms of the vowel spectral change. Some vowels such as /a/ and /u/ demonstrated similar density function of trajectory length in children and adults. This indicates that the development of vowel spectral change is nonparallel across all vowels. Children at 3 years old can possess adult-like pattern of formant trajectory on certain vowels.

The present study may bring new insights in our understanding of language acquisition. As mentioned earlier, a great deal of previous studies claimed that vowels are usually acquired before 3 years of age. Indeed, the F1 and F2 midpoint scatter plot in the present study did show separate Mandarin vowel categories in each group of child speakers. However once we scrutinize the fine structure of the dynamic acoustic features in these children’s vowel production, it can be seen that children at this age range still show considerable differences from adults for certain Mandarin vowels in various aspects.
of phonetic features. Therefore, the acoustical development of children’s vowel production is a long-term protracted process.
References


Phonological Development in Children: 18-72 months. Carbondale: Southern 
Illinois University Press.

and adults: Implications for the development of speech motor control. Journal of 
Speech and Hearing Research, 28, 8-15.

Zhongguo Yuywen (Chinese language and Writing), 5, 444-454.


Yuyanxue (Contemporary Linguistics), 1, 93-94.

linguistic units. Journal of Communication Disorders, 39, 331-349.

in children and adults. Journal of Speech, Language, and Hearing Research, 41, 
18-30.

motor coordination in childhood and adolescence. Developmental Psychobiology, 
45, 22-33.

Smith, B. L., & Gartenburg, T. E. (1984). Initial observations concerning developmental 
characteristics of labiomandibular kinematics. Journal of the Acoustical Society of 
America, 75, 1599-1605.


Press.


Addendum

Word list of Mandarin vowels used in production data collection

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Pinyin</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>da4 xiang</td>
<td>elephant</td>
</tr>
<tr>
<td></td>
<td>da4 suan</td>
<td>garlic</td>
</tr>
<tr>
<td>i</td>
<td>pi2 qiu</td>
<td>ball</td>
</tr>
<tr>
<td></td>
<td>bi2 zi</td>
<td>nose</td>
</tr>
<tr>
<td>u</td>
<td>tu4 zi</td>
<td>rabbit</td>
</tr>
<tr>
<td></td>
<td>pu2 tao</td>
<td>grape</td>
</tr>
<tr>
<td>y</td>
<td>ju2 zi</td>
<td>orange</td>
</tr>
<tr>
<td></td>
<td>yu2 tou</td>
<td>fish head</td>
</tr>
<tr>
<td>γ</td>
<td>ge1 ge</td>
<td>older brother</td>
</tr>
<tr>
<td></td>
<td>ge1 zi</td>
<td>pigeon</td>
</tr>
</tbody>
</table>
Chapter 3: Acoustic Development of Vowel Production in Native English-Speaking Children

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 6 years of age by comparing the acoustic features of children with that 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 spectral change pattern of certain vowel plotted on the basis of formant frequency values at five 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.
3.1 Introduction

Compared to many 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 been examined for a long time. In general, there are two different methods to examine 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 native adult speakers/phoneticians to determine the age of acquisition of individual speech sounds (Davis & MacNeilage, 1990; Hare, 1983; Larkins, 1983; Paschall, 1983; Stoel-Gammon & Herrinton, 1990; Templin, 1952, 1957; Wellman, Case, Mengert & Bradbury, 1931). Although these studies vary in the specific timeline of acquisition of individual vowels, they generally agree that corner vowels /ɑ, u, i/ are acquired earlier than other vowels and vowel acquisition generally complete prior to 3 years of age. However, since previous studies adopted different criteria to determine the “mastery” of each sound, for example, Wellman et al. (1931) determined that when 75% of the subjects at a certain age can produce a speech sound correctly, the sound was regarded as being acquired. Davis & MacNeilage (1990) used 60% accuracy level as a cut point to determine if the subject can produce a vowel correctly or not. Therefore, even the sound is claimed being “acquired” by a child at a particular age does not mean that it is fully and firmly developed. Some studies argued that children typically do not master adult-like features of speech until 8-years or even later (Eguchi & Hirsh, 1969; Kent, 1976; Sander, 1972; Tingley & Allen, 1975). Indeed, a more objective analysis is needed.
to capture the subtle phonetic differences which can not be manifested by narrow transcription.

Different from the phonetic transcription, acoustic analysis has been widely used to document the developmental change of acoustic features in childhood speech as a result of the lengthening of the vocal cord. (Busby & Plant, 1995; Gilbert, Robb & Chen, 1997; Kent & Murry, 1982; Lee, Potamianos & Narayan, 1999; Liberman, 1980; Perry, Ohde & Ashmead, 2001; Whiteside, Hodgson & Tapster, 2002). Although these studies adopted different methods to extract the formants and measure the frequency values due to the difficulty of formant measurement caused by the high fundamental frequency in childhood speech, collectively, these works have shown the pattern of decreasing formant frequency values and vowel duration as a function of chronological age. This trend reflects the influence of vocal tract and vocal cord lengthening on the vowel acoustic features.

Except for the decreasing formant frequency values resulted from the growth of vocal tract, children’s speech production also show another perspective of developmental change resulted from the maturation of articulators, motor control, growing vocabulary size and cognitive abilities. A large body of research has revealed that children’s articulators go through a nonlinear and nonuniform developmental pattern (Green, Moore & Reilly, 2002; Sharkey & Folkins, 1985; Smith & Gartenberg, 1984; Smith & Goffman, 1998; Smith & Zelaznik, 2004). Even adolescence still shows significant changes in speech motor control process (Walsh & Smith, 2002; Smith, 2006). The protracted development of speech motor control is manifested in both the temporal and spectral properties of children’s speech production (Nittrouer, 1993). Not just the maturation of
speech motor control, researchers also found that children’s growing vocabulary size affects their sensitivity to phonotactic probability which then influence their accuracy of speech production (Edwards, Beckman & Munson, 2004).

Taken together, although children can articulate perceivable accurate and clear speech sounds at quite early age, their speech production may still show inconsistency relative to adult speech. The present study aims to expand previous literature on the developmental change of English children’s vowel production. Of particular interest is the acoustic development of both static and dynamic vowel features in young children and the extent to which the childhood vowel features are similar to or different from adulthood speech.

Formant frequencies at the midpoint location are used to represent the acoustic properties of vowels at the relatively steady-state portion. However, it has been widely evidenced that vowels are also characterized by the spectral change which reflect their inherent properties (Harrington & Cassidy, 1994; Hillenbrand, Getty, Clark & Wheeler, 1995; Neary & Assmann, 1986). Hillenbrand et al. (1995) examined the spectral change patterns of 12 English vowels plotted on the basis of formant frequency values at 20% and 80% temporal locations over the course of vowel duration. They found that the majority of English vowels demonstrated a great amount of spectral change. In addition, the direction of formant movement also reinforced the contrast of vowel quality which might not be manifested just based on the static features. Fox and Jacewicz (2009) found that speaker’s regional dialects showed a strong effect on their acoustic measurements such as the vowel duration, trajectory length and spectral rate of change. More recently, Jacewicz, Fox & Salmons (2011a, 2011b) investigated the vowel spectral features across
different generations in northern, midland and southern American English. The results showed substantial cross-generation change in the formant movement pattern of the vowel /ɪ, ɛ, æ, ɑ/. All these studies highlighted the importance of vowel dynamic features in quantitatively and qualitatively defining vowel qualities.

Given the significance of vowel dynamic features, recently, more attention has been casted to the development of vowel dynamic features in children. As mentioned earlier, Hillenbrand et al. (1995) examined both static and dynamic features of English vowels in child speakers too. They plotted the vowel spectral change pattern of children but did not extensively compare the children’s spectral change pattern with that of adults. Assmann and Katz (2000) reported their data of the time-varying spectra change in children (3-, 5-, and 7-year-olds) and adults. This study ran statistical analysis on the formant frequencies across speakers’ age group, vowel and sample locations. The results revealed no statistical differences which suggested that children’s vowel exhibited similar vowel spectral change pattern as adults. However, this study did not present the pattern of formant movement in children for direct comparison between children and adults. More recently, Jacewicz, Fox & Salmons (2011a) investigated the vowel spectral change in older English children aged 8 to 12 years across northern, midland and southern dialect of American English. Comparison of children’s vowel dynamics pattern with that of adult speakers from corresponding dialect regions revealed that these children have comparable patterns of vowel spectral change as adults, which indicated that children at this age have acquired sociophonetic features implicated in the spectral change pattern associated with their regional dialects. Assmann, Nearey and Bharadwaj (2013) examined the pattern of vowel spectral change in children at 5 to 18 years of age from the Dallas, Texas region.
They also found that children as young as five years old already exhibited consistent vowel spectral change pattern as adults.

In general, the majority of the available studies addressed the vowel dynamic features in relatively older children. However, few of them examined the nature of formant dynamics in younger children. As suggested by previous studies, younger children demonstrated greater variability in both temporal and spectral features of speech production (Lee, Narayanan & Byrd, 2004). It remains unclear whether children at a younger age can produce speech sounds with adult-like fine structures in their acoustic properties. As stated above, childhood speech development is a complex process which is determined by physical change of their articulators as well as many other factors. Since the effect of vocal tract lengthening on children’s formant frequencies has been extensively reported, in the present study, we aim to examine the acoustic development of vowel production in relatively young English speaking children while controlling for physical maturation of these children’s vocal tracts (through the use of standard vowel normalization schemes).

In general, the purpose of the present study is two-fold:

1. To document the profile of development in both static and dynamic vowel acoustic features in native English children aged 3-6 years.

2. To examine whether children at this age range can produce vowels in an adult-like manner.
3.2 Methods

3.2.1 Speakers

The speakers participating in this study included fifteen native English-speaking children (7 girls and 8 boys) aged 3 to 6 years and six native English-speaking adults (6 females) aged 30 to 44 years. The speakers were divided into 3 age groups (AY: 3-4, 6 children, AO: 5-6, 9 children, AA: adults). All English speaking children were born and raised in central Ohio region (Columbus, Ohio). Both of their parents also speak central Ohio English. All English speaking adults were also from central Ohio and currently live in Columbus, Ohio. The six female adult speakers are mothers of the child speakers in the present study. All speakers were reported having no speech and language disorders.

3.2.2 Stimuli

The recording material (see Addendum) included 20 English monosyllabic/disyllabic words containing 10 American English vowels /i, ɪ, e, ɛ, æ, u, o, ɑ, ʌ/ (due to the merger of /ɔ/ and /ɑ/ in most dialects of American English including Ohio English (Clopper & Pisoni, 2004; Clopper, Pisoni & de Jong 2005, Labov, Ash & Boberg 2006), just one vowel /a/ was selected in the present study). Each vowel occurred in two different words. The consonants preceding the target vowels were stop consonants besides “f” in “feet”. The consonants following the target vowels were voiceless stops or fricatives. The selection of words was based on the familiarity, word frequency (Thorndike & Lorge, 1963) and picturability.
3.2.3 Procedures

The speech samples were collected through a visual-auditory word repetition task under the control of a custom MATLAB program. The experiment included two blocks of recording session for each speaker. In each block, randomly ordered\(^6\) pictures containing target words were presented on the computer screen. Each speaker was seated in front of a laptop computer in a quite room and repeat each target word right after the audio prompt produced by a female native English speaker to a Shure SM10A head-mounted microphone situated approximately 1-inch from the speakers’ mouth. All speech productions were recorded and digitized directly onto a hard drive disk at a 16-bit quantization rate and 44.1 kHz sampling rate.

3.2.4 Acoustic measurements

3.2.4.1 Formant frequencies

Prior to acoustic analysis, all tokens were down-sampled to 11 kHz using a custom MATLAB program. Some studies (Hillenbrand, et al., 1995; Neary & Assmann, 1986; Assmann & Katz, 2000) adopted two measurement points over the vowel duration to define the dynamic formant movement, one at the onset location and one at the offset location. However, it has been shown that formant movements are not always a straight line for certain vowels. To better estimate the curved formant track and reduce the redundant calculation caused by more dense sampling points, formant frequencies are measured at five equidistant temporal locations (20-35-50-65-80% point) using

\(^6\) The same random order was used for both recording blocks.
spectrographic analysis program TF32 (Milenkovic, 2003), consistent with previous studies (Fox & Jacewicz, 2009; Jacewicz, Fox & Salmons, 2011b; Jacewicz, Salmons & Fox, 2009). The landmark locations of vowel onset and offset were located by hand on the basis of the waveform as well as visual check of the spectrogram. Specifically, 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 at which the amplitude of the waveform dropped significantly due to the following stop closure or frication.

3.2.4.2 Vowel normalization

A set of normalized formant frequency values was generated using the method in Lobanov (1971) to avoid the effect of different size of vocal tract among speakers and lengthening of vocal tract as a function of chronological age on the formant frequency values. Since the normalized formant frequency values do not reflect Hz values, they were then rescaled into Hz-like values using the method in Thomas and Kendall (2007) to facilitate interpretation of the normalized vowel spaces.

3.2.4.3 Trajectory length

Trajectory length (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%), was calculated on the basis of the rescaled normalized F1 and F2 values in the acoustic vowel space (Fox & Jacewicz, 2009) using the following formula:
\[ TL = \sum_{n=1}^{4} VSL_n \]  \hspace{1cm} (3)

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

\[ VSL_n = \sqrt{(F_{1n+1} - F_{1n})^2 + (F_{2n+1} - F_{2n})^2} \]  \hspace{1cm} (4)

The measure of trajectory length provides a detailed assessment of the magnitude of formant movement over the course of vowel duration between the 20 to 80% points especially for the curved formant trajectory which has formant frequency values in later temporal portions bend over to that in the former temporal portions. In specific, a longer trajectory length indicates the greater change of formant frequency values and greater magnitude of formant movement.

3.3 Results

3.3.1 Vowel duration

Vowel duration was calculated on the basis of the manually located onset and offset landmarks (shown in Figure 3.1 and Table 3.1). Previous literature has revealed that different vowels in different dialects of American English are characterized by distinctive temporal features (Jacewicz, Fox & Salmons, 2007). Thus a significant effect of vowel on duration was anticipated and was not of particular interest in the present study. One-way ANOVA and Tukey HSD were used to examine the effect of age on vowel duration for each English vowel (results summarized in Table 3.1). Generally, child speakers produced longer vowel duration than adult speakers even though the decreasing pattern of vowel duration from children to adults was not statistically significant. In addition, young children showed relatively larger variation in the vowel
duration than did older children and adults for most English vowels. This suggested that young children were less consistent in terms of the temporal properties of vowel production than were older children and adults.

Figure 3.1. Bar plots showing the mean and standard error of vowel duration for each English vowel in each age group.
Table 3.1. Summary of statistical results showing the effect of age on vowel duration of each English vowel (means and standard errors are in ms).

<table>
<thead>
<tr>
<th></th>
<th>AY M (SE)</th>
<th>AO M (SE)</th>
<th>AA M (SE)</th>
<th>P</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>174 (19)</td>
<td>186 (10)</td>
<td>168 (11)</td>
<td>0.596</td>
<td>0.056</td>
</tr>
<tr>
<td>ï</td>
<td>111 (11)</td>
<td>119 (8)</td>
<td>91 (5)</td>
<td>0.088</td>
<td>0.237</td>
</tr>
<tr>
<td>e</td>
<td>174 (20)</td>
<td>179 (13)</td>
<td>154 (12)</td>
<td>0.502</td>
<td>0.074</td>
</tr>
<tr>
<td>æ</td>
<td>136 (13)</td>
<td>147 (9)</td>
<td>133 (12)</td>
<td>0.617</td>
<td>0.052</td>
</tr>
<tr>
<td>u</td>
<td>151 (16)</td>
<td>182 (12)</td>
<td>190 (21)</td>
<td>0.261</td>
<td>0.139</td>
</tr>
<tr>
<td>ð</td>
<td>146 (16)</td>
<td>180 (11)</td>
<td>169 (10)</td>
<td>0.172</td>
<td>0.178</td>
</tr>
<tr>
<td>o</td>
<td>101 (10)</td>
<td>110 (7)</td>
<td>88 (6)</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>ø</td>
<td>163 (24)</td>
<td>178 (10)</td>
<td>176 (16)</td>
<td>0.799</td>
<td>0.025</td>
</tr>
<tr>
<td>a</td>
<td>153 (21)</td>
<td>166 (14)</td>
<td>167 (14)</td>
<td>0.802</td>
<td>0.024</td>
</tr>
<tr>
<td>æ</td>
<td>118 (13)</td>
<td>129 (14)</td>
<td>114 (11)</td>
<td>0.726</td>
<td>0.035</td>
</tr>
</tbody>
</table>

3.3.2 Midpoint F1 by F2 vowel space

Based on the rescaled normalized F1 by F2 at midpoint location, the scatter plots of English vowel categories were presented in Figure 3.2. Following the method used in Zhou and Xu (2008), vowel ellipse was plotted for each vowel category to encircle approximately 95% samples (ellipse area for each vowel in each age group is shown in Table 3.2). Specifically, a linear fit was first adopted to determine the rotation degree and the direction of the ellipse semimajor. Then the ellipse center was set at the mean of rescaled normalized F2 along the fitting line and the length of semimajor axis was set to two standard deviations of the data points away from the center. The semiminor axis was set at the direction perpendicular to the semimajor axis and its length was set to two standard deviations of the data points away from the center. As can be seen, younger children produced these English vowels with greater variation in terms of the positional
distribution of each vowel category. Some vowel categories were substantially overlapped, which indicated the less definable acoustic vowel categories in younger children. For older children, the sizes of ellipses for most vowel categories especially back vowels substantially decreased, which resulted in the reduced overlaps among these vowels.

From younger children to older children, the front vowels did not show much observable change in terms of the positional distribution of each vowel categories. However, the back vowels displayed considerable change particularly in the vowels /u/, /o/ and /o/. For adult speakers, these ten vowels were well separated and individual samples in each vowel category were tightly concentrated. Compared with child speakers, the biggest change resided in the complete separation of high front vowels /i/, /ɪ/ and /e/ and high back vowels /u/ and /ʊ/. In addition, the size of ellipse for each acoustic vowel category also reduced considerably from children to adults. The continuing decreasing ellipse area implied the reduced variation of vowel production from children to adults. One thing need to be pointed out was that although the vowel /u/ showed 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. As mentioned earlier, all native English speakers were from central Ohio. One of the important phonetic features of midland English is the tendency of fronting /u/. The variation in F2 of /u/ in the present study evidenced the pattern of sound change in the vowel /u/ in Midland English.
Figure 3.2. Scatter plot of midpoint formant frequency values for individual English vowel categories in each age group.
Table 3.2. The ellipse areas of individual English vowels in each age group (in Hz^2).

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

Figure 3.3 shows the means and standard errors of midpoint formant frequency values for each vowel in each group of speakers. It can be observed that both groups of children deviated from adult norms for some English vowels. In order to further examine to what extent the relative position of each English vowel changes as a function of age, one-way ANOVA was used separately on rescaled normalized F1 and F2 across three age groups for each vowel. In the front vowels, there was a significant age difference on rescaled F2 for the vowel /e/ (F(2,18) = 11.945, p < 0.001, η_p^2 = 0.570) and /ɛ/ (F(2,18) = 4.718, p = 0.023, η_p^2 = 0.344). The results of Tukey HSD showed that both younger and older children produced significantly lower F2 than did adults for the vowel /e/ while only older children produced significantly higher F2 than did adults for the vowel /ɛ/. In the back vowels, there was a significant age effect of rescaled F1 for the vowel /o/ (F(2,18) = 4.466, p = 0.027, η_p^2 = 0.332). Tukey HSD revealed that younger children produced significantly higher F1 than did adults. For the vowel /ʌ/, one-way ANOVA results showed a marginal significant effect of age on the F2 (F(2,18) = 3.414, p = 0.055, η_p^2 = 0.275). Tukey HSD showed that younger children produced significantly higher F2 than did older children.
3.3.3 Vowel space area

Vowel space area has been widely used as an important index to signify the development of children’s vowel system as a function of vocal tract lengthening. Previous studies have also shown that the size of working vowel space area is an important index for speech intelligibility and communication disorders. In English, the size of working vowel space is determined by the four corner vowels /i, æ, a, u/. Voperian & Kent (2007) found that English vowel space area shows a general decreasing pattern as a function of chronological age as a result of the decreasing formant frequency values. However, few studies have examined the extent to which the children’s working vowel space area is similar to or different from that of adults after eliminating the
developmental change of vocal tract length. If the normalized vowel space areas of children are similar to those of adults, it suggests that children at this age have developed the same basic framework of the vowel system as adults. Otherwise, it indicates that children at this age are still developing these boundary vowels and haven’t developed adult-like vowel frames. In the present study, the vowel space area of each speaker for all three age groups was calculated on the basis of rescaled normalized formant frequency values of the four English corner vowels (shown in Figure 3.4). The one-way ANOVA result revealed no significant difference on the normalized vowel space area among these three groups of speakers ($F(2,18) = 0.484, p > 0.05, \eta^2_p = 0.051$), which indicated that children at this age had had developed an adult-like vowel space.

![Figure 3.4](image)

Figure 3.4. Box plots of English vowel space area in each age group. The vowel space area of individual speaker within each group was also demonstrated. Each box depicts the lower quartile, median and upper quartile. Each asterisk represents a data point of one subject.
3.3.4 Formant dynamics

3.3.4.1 Formant movement pattern

Figure 3.5 shows the formant trajectories (in the F1 x F2 plane) for these ten English vowels in each age group. For adult speakers, these ten vowels demonstrated different patterns of vowel spectral change in terms of the position, direction and magnitude of the formant trajectories. In particular, /e/ and /o/ showed the most extensive formant movement. The substantial formant movement of these two vowels demonstrated their nature of diphthongization. For the other vowels, in the serial of front vowels, /i/ remained relatively stable in F1 and showed little increase in F2. /ɪ/ again, maintained relatively stable in F1 but showed some decrease in F2. /ɛ/ and /æ/ showed similar formant trajectories represented by increase in F1 and little decrease in F2. But these two vowels were distinctive in their relative positions in the F1 by F2 space. In the series of back vowels, /u/ showed a decrease in F2 and little change in F1. The vowel /ʊ/ showed slight change in F1 but no obvious change in F2. The vowel /ʌ/ showed a bowed formant movement due to a reversal of direction in F1 and a decrease in F2. The vowel /a/ showed moderate increase in F1 but little change in F2.

As for children, the general shape of formant trajectory in each vowel was preserved to a large extent. However, differences in the patterns of formant trajectories and relative positions can still be observed in most vowels. For example, /e/ and /o/ were
produced in a more fronted position as a function of speakers’ age. /i/ showed a
downward movement from children to adults and /a/ demonstrated a forward and slightly
downward movement. The vowel /ʌ/ also showed position difference from children to
adults. Except for the relative positions, child speakers also demonstrated different shapes
of formant trajectories in terms of the direction and magnitude of formant trajectories for
certain vowels such as /u, ε, æ, ʌ/. For example, the formant trajectory of /u/ in young
children showed slight increase in F2 and very little increase in F1, which completely
differed from that of adults. The formant trajectories of /ε/, /æ/ and /ʌ/ were more curved
in adults than in both groups of children.

Figure 3.5. Vowel dynamic 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 80% point.
3.3.4.2 Trajectory length

To further examine the extent to which the magnitudes of formant movement differ across three groups of children, the lengths of formant trajectories were calculated on the basis of rescaled normalized formant frequency values for each vowel in each age group (shown in Figure 3.6). The trajectory lengths increased from children to adults for most vowels /ɪ, e, ɛ, o, ɑ, ʌ/. But for certain vowels like /i, u, ʊ/, the trajectory length show slightly decrease or no apparent change from children to adults. One-way ANOVA was used to test how the age factor influences the trajectory length for each vowel. Tukey HSD was then used to examine the nature of significant differences between the age groups (summarized in Table 3.3). The results show that for the vowel /æ, ʌ/, adults have significantly longer trajectory length than both younger and older children. For the vowel /ɪ, ɑ/, adults have significantly longer trajectory then than older children. For the vowel /e/, one-way ANOVA result showed significant effect of age, but Tukey HSD tests just showed marginal significance between adults and both younger and older children.
Figure 3.6. Bar plots showing the means and standard errors of trajectory length for each English vowel in each age group.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>AY M (SE)</th>
<th>AO M (SE)</th>
<th>AA M (SE)</th>
<th>P</th>
<th>η²</th>
<th>Significant difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>134 (20)</td>
<td>101 (8)</td>
<td>105 (7)</td>
<td>0.147</td>
<td>0.191</td>
<td></td>
</tr>
<tr>
<td>ì</td>
<td>91 (12)</td>
<td>74 (2)</td>
<td>113 (13)</td>
<td>0.021</td>
<td>0.35</td>
<td>AO&lt;AA</td>
</tr>
<tr>
<td>e</td>
<td>190 (20)</td>
<td>181 (21)</td>
<td>194 (16)</td>
<td>0.898</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>é</td>
<td>98 (9)</td>
<td>101 (4)</td>
<td>132 (15)</td>
<td>0.039</td>
<td>0.303</td>
<td>AY&lt;AA, AO&lt;AA</td>
</tr>
<tr>
<td>æ</td>
<td>90 (7)</td>
<td>105 (6)</td>
<td>150 (22)</td>
<td>0.012</td>
<td>0.39</td>
<td>AY&lt;AA, AO&lt;AA</td>
</tr>
<tr>
<td>u</td>
<td>134 (20)</td>
<td>199 (37)</td>
<td>137 (19)</td>
<td>0.237</td>
<td>0.148</td>
<td></td>
</tr>
<tr>
<td>ø</td>
<td>87 (10)</td>
<td>82 (7)</td>
<td>84 (7)</td>
<td>0.883</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>191 (22)</td>
<td>188 (15)</td>
<td>225 (12)</td>
<td>0.249</td>
<td>0.143</td>
<td></td>
</tr>
<tr>
<td>ʌ</td>
<td>125 (8)</td>
<td>101 (5)</td>
<td>135 (11)</td>
<td>0.014</td>
<td>0.377</td>
<td>AO&lt;AA</td>
</tr>
<tr>
<td>ʌ</td>
<td>102 (6)</td>
<td>109 (7)</td>
<td>150 (10)</td>
<td>0.002</td>
<td>0.514</td>
<td>AY&lt;AA, AO&lt;AA</td>
</tr>
</tbody>
</table>

Table 3.3. Summary of statistical results showing the effect of age on trajectory length of each English vowel (means and standard errors are in Hz).
To visually present the distribution pattern of trajectory lengths in children and adults, kernel density estimation was used to generate the smooth distribution curves of trajectory length of three age groups for each English vowel (shown in Figure 3.7). In general, kernel density estimation generates a smooth probability density function on the basis of observed data samples to represent the distribution of the unknown target population. This method is closely associated with histogram but endowed with the characteristics of continuity by selecting an appropriate parameter of smoothness. In each distribution curve, the location of the peaks indicates where the highest density of the variable is encountered and the spread indicates the variability of the distribution.

As shown in this figure, the distribution curves of the vowels /i, e, u, o, ʌ/ in children showed larger variability than those in adults. In addition, the curves of the vowels /e/, /o/, /ʌ/ showed mismatched peak locations relative to those of adults. The different patterns of distribution curves between children and adults indicate that children at this age had not developed an adult-like manner in terms of the magnitude of the formant movement. The distribution of /ʊ/ in children showed similar spread and peak location to those of adults. This indicates that children produced this vowel with similar magnitude of formant movement to those of adults. The distribution curves for the vowels /ɪ/, /ɛ/, /æ/ and /ɑ/ in adults showed greater variability than those in children. This indicates that adults produced these vowels in a more variable manner in terms of their magnitude of formant movement while children were more consistent in the formant movement feature. The larger variation in adult’s production of formant movement of
these vowels may be associated with the vowel change in American English. Northern Cities Vowel Shift includes a series of vowel changes involved with rising and fronting of /æ/, backing of /ɛ/ and /ʌ/, lowering and fronting of /a/, lowering and backing of /i/ etc. (Clopper, Pisoni & de Jong, 2005; Jacewicz, Fox & Salmons, 2011c). Previous literature suggested that young adults especially young adult women generally lead the vowel change (Jacewicz, Fox & Salmons, 2006). In the present study, all adult speakers were the mothers of the child speakers and aged in mid 30s. The vowel changes may be present and active in these adult speakers and thus caused the greater variation in the formant dynamic features of the involved vowels in these adult speakers relative to the child speakers.

Figure 3.7. Probability density function of trajectory length for each Mandarin vowel across three age groups.
Figure 3.7 continued
Figure 3.7 continued
Figure 3.7 continued
In addition to the developmental change from children to adults, it should be noted that some vowels demonstrated skewed distribution curve of trajectory lengths in some age groups. As can be seen in Table 3.4, the vowels /i/, /ɪ/, /u/ and /o/ show skewed distributions for the younger children. The vowels /i/, /ɪ/ and /e/ have skewed distributions for the older children and the vowels /ɪ/, /e/ and /u/ show skewed distributions for the adults. In addition, some vowels such as /i/, /ɪ/, /e/, /u/, /o/ and /ʌ/ have bimodal distributions for both children and adults. Therefore, the density function analyses provide important information which may not be evident from means and standard errors and may not be appropriately shown by ANOVA test.

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>ɪ</th>
<th>e</th>
<th>ɛ</th>
<th>æ</th>
<th>u</th>
<th>ʊ</th>
<th>o</th>
<th>ʌ</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY</td>
<td>1.193</td>
<td>1.065</td>
<td>0.400</td>
<td>0.158</td>
<td>0.581</td>
<td>0.722</td>
<td>1.539</td>
<td>1.184</td>
<td>0.441</td>
</tr>
<tr>
<td>AO</td>
<td>1.215</td>
<td>1.105</td>
<td>1.833</td>
<td>-0.081</td>
<td>0.257</td>
<td>0.954</td>
<td>0.221</td>
<td>0.591</td>
<td>0.690</td>
</tr>
<tr>
<td>AA</td>
<td>0.262</td>
<td>0.957</td>
<td>-0.977</td>
<td>0.384</td>
<td>0.516</td>
<td>1.324</td>
<td>0.081</td>
<td>-0.214</td>
<td>-0.060</td>
</tr>
</tbody>
</table>

Table 3.4. Skewness of the distribution curves of trajectory length for individual English vowels in each age group.

### 3.4 Discussion and Summary

This study examined the development of acoustic properties in vowel production by native English children aged 3 to 6 years. Of particular interest is how children at this age are similar or different from native English adults in both static and dynamic vowel features. By means of vowel normalization, we excluded the effect of increasing vocal tract size on formant frequency values. The results demonstrated that children at this age
range still show developmental change in their organization of acoustic vowel categories and did show different pattern 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, these ten English vowel categories generally exhibited a decreasing trend of overlap in the F1 by F2 vowel space as a function of age. Unlike well separated vowel categories in the F1 x F2 space, 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 greatly while the front vowels remained relatively stable. Compared with the organization of the vowel categories in adult’s acoustic space, the relatively stable front vowels while reduced overlap in the back vowels from younger children to older children group suggested an inconsistency in the development of the front and back vowels.

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 (Green et al. 2002; Walsh & Smith, 2002). Most of back vowels in English are round vowels which are produced with lips protruded or rounded. As a result of the protracted development of motor skills related to the lips, the acoustic manifestation demonstrated greater variability. Except for the back vowels, front vowels also demonstrated reorganization from older children to adults. The significant reduction in the spread of /i/ and /e/ in the acoustic vowel space made the three high front
vowels /i, ɪ, e/ completely separated. The continuing development of front vowels from older children to adults suggested that even though the acoustic development of the front vowels may occur at relatively young age, the completion of establishment of vowel categories is a long-term and complex process.

Except for the static acoustic features, vowel dynamics have also been investigated in the present study. The present study used formant frequencies at five time locations to calculate the trajectory length and plot the formant tracks. Generally, both groups of children produced comparable shape of formant movement as adults. However, the formant tracks produced by children do not perfectly match with those produced by adults in either the relative position or the length of the formant trajectories. The vowels /e, ɪ, o, ʌ, a/ showed apparent positional change from children to adults. For example, the vowel /ɪ/ moved downward and the vowel /a/ showed clockwise movement from children to adults. These positional changes are consistent with the finding of Jacewicz, Fox & Salmons (2011c), which provide extra evidence for the mechanism of cross-generational vowel change in Ohio English.

In addition to the positional change, children’s vowel production also differs from adults in the magnitude of formant movement. In general, children vowel production is characterized by shorter trajectory length. In addition, as shown in the density plot of trajectory length for each vowel across three age groups, the distribution curve of the trajectory length in children clearly demonstrates developmental change as a function of age for certain vowels such as /i/, /e/, /u/, /ʌ/, /o/ and /a/. The greater spread of TL distribution curve in children for these vowels indicates that children are less consistent
in terms of the magnitude of spectral change. The acoustic measurements of spectral change in children still differ from adult targets.

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 growth in vocal tract. The acoustical development of vowels from children to adult norms is a long-term and complicated process.
References


**Addendum**

Word list of English vowels used in production data collection

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>feet</td>
</tr>
<tr>
<td>i</td>
<td>kiss</td>
</tr>
<tr>
<td>e</td>
<td>cake</td>
</tr>
<tr>
<td>e</td>
<td>desk</td>
</tr>
<tr>
<td>æ</td>
<td>cat</td>
</tr>
<tr>
<td>u</td>
<td>boot</td>
</tr>
<tr>
<td>o</td>
<td>book</td>
</tr>
<tr>
<td>o</td>
<td>coat</td>
</tr>
<tr>
<td>a</td>
<td>box</td>
</tr>
<tr>
<td>ʌ</td>
<td>cup</td>
</tr>
<tr>
<td>ʊ</td>
<td>geese</td>
</tr>
<tr>
<td>ɪ</td>
<td>kitchen</td>
</tr>
<tr>
<td>ɪ</td>
<td>tape</td>
</tr>
<tr>
<td>ɛ</td>
<td>pet</td>
</tr>
<tr>
<td>æ</td>
<td>bat</td>
</tr>
<tr>
<td>u</td>
<td>goose</td>
</tr>
<tr>
<td>o</td>
<td>cookie</td>
</tr>
<tr>
<td>o</td>
<td>boat</td>
</tr>
<tr>
<td>a</td>
<td>stop</td>
</tr>
<tr>
<td>ʌ</td>
<td>duck</td>
</tr>
</tbody>
</table>
Chapter 4: Vowel Development in an Emergent Mandarin-English Bilingual Child: A Longitudinal Study

Abstract

This longitudinal case study documents the emergence of bilingualism in a young boy on the basis of a detailed acoustic phonetic analysis of his vowel productions. The child was raised as monolingual Mandarin (L1) until his enrollment in an all-English (L2) preschool at the age of 3;7. Following enrollment, he was recorded via a picture-naming task on a regular basis over 20 months. The study examined (1) his initial L2 vowel space, (2) the process of L1-L2 separation, and (3) his L1 vowel system in relation to L2. The child initially utilized his L1 base in building the L2 vowel system. The L1-L2 separation began by means of a drastic restructuring of his working vowel space to create maximal contrast between the two languages. This abrupt partitioning was accomplished by temporarily creating a reduced L2 vowel space, which gradually expanded as the child “added” L2 vowels to his L2 system by producing a greater number of acoustic distinctions. While the general shape of his L1 vowel space remained unchanged throughout, L1 developmental processes and influence of L2 were also in effect. The developmental profile of this child uncovered strategies sequential bilingual children may use to restructure their phonetic space and construct a new system of contrasts in L2.
4.1 Introduction

Emergent bilingualism is a language contact situation faced by children who learn their first language (L1) from birth as the result of linguistic input in their L1-speaking home environments and develop their second language (L2) later in childhood from input received from playmates and the school (Verhoeven, 2007). The interest in language development in emergent bilinguals has increased in recent years, as the changing global demographics and population mobility have impacted young children’s language acquisition in the social context of immigration. However, instrumental studies which examine L2 development in these sequential bilinguals are sparse and experimental work is primarily focused on language development in simultaneous bilinguals. The goal of the present study is to provide new experimental evidence pertaining to how emergent bilinguals separate the two languages and, in so doing, contribute to the long-standing debate on language separation (“one system or two”) in bilingual children.

The existing research reporting longitudinal data are mixed. On one hand, it has been shown that, in sequential bilinguals, the L1 phonological system initially dominates the L2 and serves as a foundation for subsequent development and eventual separation of L2 (Fantini, 1985; Watson, 1991). On the other hand, it has been demonstrated that children can separate the two phonological systems from the beginning (Holm & Dodd, 1999). This was evident in both children’s correct productions of language-specific phonemes and in the types of errors they demonstrated, which conformed to phonotactic rules in each language.

To date, due to the paucity of experimental data, little is known about phonetic development and phonetic separation of the two sound systems in sequential bilingual
children. The present longitudinal case study aims to contribute to a better understanding of this process by documenting phonetic development in a young boy who participated in this research over a period of 20 months. The child was born in the United States to Mandarin-speaking parents and was raised as monolingual until the age of 3;7. To determine the developmental profile of his vowel system(s), a detailed acoustic phonetic analysis of his vowel productions in Mandarin and English was conducted. Our research interests are three-fold. First, we aim to determine the initial state of his L2 vowel space. Second, we examine the process of language separation during his subsequent exposure to English in the preschool. Finally, we aim to establish whether and how his L1 Mandarin system has changed as a function of his L2 development.

An acoustic study by Simon (2010) is of immediate relevance to our work. That study examined the acquisition of English (L2) voice contrast in word-initial stops by a native Dutch (L1) child. Both English and Dutch have a voiced-voiceless stop contrast. However, while the English contrast is represented as short-lag (voiced) versus long-lag (voiceless) distinction, the Dutch contrast is manifested differently, as prevoiced (voiced) versus short-lag (voiceless) distinction. The child successfully mastered the English contrast within a 7-month observational period. Meanwhile, his Dutch voice contrast boundary shifted toward that in English as a result of the influence of his L2, indicating that the L1 phonetic system could still be restructured as a function of L2 acquisition.

Bilingual vowel systems have been studied primarily in adults and adolescents (rather than in children) who learned their L2 later in childhood or even in early adulthood. Acoustic analysis was used to uncover the production differences such as between groups of L2 speakers differing in the amount of experience or between L2
speakers and the monolingual speakers. Several findings from this research are of importance to our study. First, Guion (2003) showed that adult Quichua-Spanish bilinguals who were immersed in L2 earlier in life were more successful in learning the new vowel contrasts in L2 than those who were immersed in L2 later in life. However, vowel productions of all these sequential bilinguals—regardless of their success with the acquisition of contrasts—were still different from those of the monolingual speakers. In addition, the fine-grained level of acoustic detail also showed that the L2 vowel system could also introduce changes to the L1 system.

Second, Baker and Trofimovich (2005) demonstrated how L1 (Korean) and L2 (English) vowel systems interacted in bilinguals who differed in terms of the age of exposure to L2 (older children or young adults) and the amount of experience with L2 (1 or 7 years). Of interest to our study is the way these bilinguals organized a subset of L2 vowels. For the adults, there was a unidirectional influence of the L1 on the L2, which was primarily determined by cross-language similarity of L1 and L2 categories. In the process of L2 learning, their L2 vowel system underwent restructuring but this L2 restructuring did not affect their L1 system. In contrast, in older children, phonetic restructuring as a function of L2 learning affected both languages, particularly for those who had used their L2 for seven years. Thus, these child bilinguals with extended L2 use produced L1 and L2 vowels that were more demonstrative of bidirectional influences. These results suggest that L1 and L2 interact differently depending on the age at which L2 is learned.

In the present study, the Mandarin-speaking child was exposed to a much richer English vowel system. Specifically, English has at least 12 monophthongal vowels (/i, ɪ,
e, ɛ, æ, u, o, ʌ, ɔ, ɑ, ɚ/) while Mandarin has only five basic monophthongal vowel phonemes (/a, i, u, y, ɤ/). In addition, the English vowel system has a tense-lax distinction in several vowel pairs while the Mandarin system lacks this contrast. Based on the research reviewed above, there are several possibilities as to the developmental paths of the child’s L1 and L2. With respect to his L1 Mandarin vowel system, there are at least two possible outcomes. First, the L1 system remains intact over the course of L2 development if the child has indeed acquired it. Second, if the child has not yet fully developed his L1 vowel system, some of his L1 vowels may still be “flexible” and susceptible to restructuring as a function of category formation in L2. Consequently, changes to his L1 may still be expected, at least for selected vowels.

In terms of his L2 English, our working hypothesis is that the child will begin constructing his L2 vowel space on the basis of his acquired L1 space. Because of his young age, this unidirectional influence of L1 on L2 is expected to decrease systematically with his increased experience with L2 so that his English vowel system will eventually diverge from Mandarin. Tentatively, we predict that the child will construct his L2 vowel system by first creating L1-based categories for acoustically similar L1 and L2 vowels (Mack, 1990). Then, over the course of his L2 development, these clustered categories should separate into individual L2 vowels as the child gradually develops a distinct L2 system in relation to L1. This developmental category separation as a function of language experience has been well documented in monolingual infants (Kuhl & Meltzoff, 1996) and we expect the same process to occur in L2 acquisition (e.g., Baker & Trofimovich, 2005; Flege, 1995; Guion, 2003).
4.2 Methods

4.2.1 Participant

One native Mandarin-speaking male child participated in this longitudinal study. He was born and raised in the United States. Both of his parents were native Mandarin speakers who immigrated to the United States one and a half years before he was born. His family moved from Chicago to Columbus Ohio in July, 2009 when he was 2;10. From birth, he was immersed in a Mandarin-dominant environment until his enrollment in an English preschool program at the age of 3;7. After that, he received substantial exposure to English as he began to learn it as his L2. The first recording took place two weeks after he started the preschool program at the age of 3;8. His parents reported that he had no hearing or speech impairment.

4.2.2 Language context

The child received his L1 Mandarin input primarily from his parents, predominantly his mother. Both parents received at least a college-level education in China and spoke Mandarin with each other in their daily life. They also used Mandarin to interact with the child. Besides interacting with his family members, the child played mainly with Mandarin-speaking children and had a very limited contact with English. After enrolling in the English preschool, he was immersed in an all-English L2 environment three days per week. All three preschool teachers were native English speakers of central Ohio dialect. Only one out of the other 24 children in his class was non-native English speaker but he also spoke English with the participant in this study.
All class instructions and materials were in English. When he was 4;11, he was enrolled in a full-time (5 days per week) kindergarten program. Both of his kindergarten teachers were native English speakers of central Ohio dialect and he was the only non-native English student in the whole class totaling 16 children. The child’s parents reported that he preferred to use English at home after starting the kindergarten program, indicating that English was gradually becoming his dominant language.

### 4.2.3 Speech material

The recording materials included two sets of picturable words. One set represented Mandarin monosyllabic and disyllabic words each of which included one of five basic Mandarin monophthongal vowel phonemes: /i, y, a, u, ɤ/. In a widely accepted opinion concerning the nature of Mandarin vowel system, Lin and Wang (2001) identified eight monophthongal vowels (/i, ɿ, ʅ, y, u, a, ɤ, o/) in the Mandarin vowel system. Among these eight vowels, /ɿ/ and /ɿ/ are allophonic variations of the vowel /i/.

The vowel /o/ is often considered a diphthong rather than a monophthong (Zee, 2001). Therefore, the present study follows Duanmu (2000) which proposed that Mandarin contains five basic vowel phonemes /a, i, u, ɿ, ə/. Here we use the symbol /ɤ/ rather than /ə/ to refer to the fifth vowel as it is a more standard usage. The other set represented monosyllabic English words each of which included one of 11 basic monophthongal American English vowels: /i, ɪ, e, æ, u, ʊ, o, ɔ, ɒ, ʌ/. All words were familiar to young children and beginning learners. The basis for word selection included familiarity, word
frequency (Thorndike & Lorge, 1963) and picturability—phonetic context was not strictly controlled in either the Mandarin or English word sets (although all vowels were produced in a stressed syllable), nor was the tone environment controlled for in the Mandarin word set. In each recording session, the child produced each token once for a total of 19 tokens in the Mandarin word set and 33 tokens in the English word set (these target words used to elicit their productions are shown in the Addendum).

4.2.4 Procedures

The study extended over a 20-month period. The recording procedure included two phases: In the first 12 months, one recording session was conducted each month; the average time between sessions was 31 days. After the child had relatively well-established acoustic vowel spaces in both Mandarin and English, recordings were made in months 15, 16, 19 and 20. The word productions were recorded through picture-naming task in a quiet room at the child’s home with his mother present. In each session the child was first recorded saying the Mandarin words and then, after a short 15-20 minute break, the English words. The same experimenter (a fluent bilingual Mandarin-English speaker) used Mandarin to interact with him in the Mandarin sessions and English in the English sessions. During these sessions, the child was seated in front of a laptop computer wearing a Shure SM10A head-mounted microphone situated approximately one inch from his mouth. Pictures representing target words were randomly ordered and presented on the computer monitor (the same random order was used across all recording sessions). Each target word was produced once and was
recorded directly onto a hard drive disk with a 16-bit quantization rate and 44.1 kHz sampling rate. All recordings were done under the control of a custom Matlab program.

### 4.2.5 Acoustic measurements

Spectrographic analysis was used to determine the frequencies of the first two formants, F1 and F2. The formant frequencies were measured at the vowel’s midpoint, which was determined on the basis of temporal locations of each vowel’s onset and offset in the waveform using a custom Matlab program. A second speech analysis program TF32 (Milenkovic, 2003) was used for an additional visual check of the spectrograms and for hand correction of the automatic formant measurements, if needed. Given the child’s relatively high F0, a large analysis bandwidth (600Hz) was used. An auditory check of the vowel quality was also done to ensure that no part of a preceding or following consonant was included (this was important for vowels that were preceded or followed by the sonorants /r, l, w/).

Vowel onsets and offsets were determined using standard measurement criteria (Kent & Read, 1992). Vowel onset following the oral stops /b, p, pʰ, t, d, k/ was defined as the initial zero crossing of the first period of voicing following stop closure release. Vowel onset following the fricatives /f, s, ʃ, x, h/ and the affricates /tʃ, tʃʰ/ was defined as the zero crossing point of the first period of voicing following cessation of frication. Vowel onset following the nasal stops /m, n/ was defined as at the point of significant increase in amplitude of the waveform; this was co-located with a visual check of the end of the nasal murmur in the spectrogram. Vowel onset following the
lateral /l/ was set at the point of significant increase in amplitude of the waveform with visual check of discontinuity of formant transitions in the spectrogram due to the release of tongue constriction of the /l/. Vowel onset following the rhotic /r/ was defined as the point of significant increase in amplitude of the waveform with visual check of the rise of F3 to a stable pattern for the following vowel. Vowel onset following the glide /w/ was set at point of significant increase in amplitude of the waveform with visual check of the attainment of relative stable pattern of F2 of the following vowel.

Vowel offset preceding oral stops /p, b, t, d, k, g/ was set at the zero crossing point closest to the significant decrease in amplitude due to the stop closure. Vowel offset preceding fricatives was marked at the point prior to the onset of frication. Vowel offset preceding nasal /n/ was set at the onset of nasal murmur represented by significant decrease in amplitude. Vowel offset preceding lateral /l/ was located at the end of relatively steady state of the target vowel associated with visual check of the start point of weak energy in relatively mid to high frequency region due to lateral zero and auditory check of any “l” quality in the vowel.

4.3 Results

Preliminary analyses demonstrated that the child’s productions did not change considerably within each two-month period. For that reason, the data were collapsed into two-month epochs and the results are presented for each epoch in lieu of each individual session. Results are first presented for basic dispersion patterns of Mandarin and English vowels in their respective vowel spaces followed by an analysis of the vowel space areas.
Figure 4.1. Mean formant frequency values (with standard error) for L1 Mandarin (in black) and L2 English (in red) vowels produced by the child across eight epochs. Lines connect the three vowels /i, a, u/ in Mandarin (forming a triangular vowel space), the three vowels /ɪ, ʊ, ʌ/ in English in epoch 1 (a triangular vowel space similar to Mandarin) and the four corner vowels /i, æ, a, u/ in English from epoch 2 to epoch 8 (forming a quadrilateral space).
4.3.1 Organization of vowel systems in L1 Mandarin and L2 English

Shown in Figure 4.1 are the relative positions of L1 Mandarin and L2 English vowels over the eight epochs, superimposed in the common F1 x F2 plane. Unnormalized (rather than normalized) formant frequency values were used in the plots because normalization affects the overall size of the vowel space but not the relative position of individual vowels. It is the structure of the vowel system in the acoustic space in terms of the relative position of vowels – which is not changed by vowel normalization – that is of interest here. As can be seen, the child’s English vowel space changed over the 20-month period from a Mandarin-like triangular shape to a typical English-like quadrilateral shape. The development of his L2 vowel system can be divided into three stages: Initiation Phase, Reorganization Phase and Stabilization Phase.

In the Initialization Phase (epoch 1), the basic shape of the English vowel system—the /i, u, æ, ɑ/ quadrilateral—was absent while the vowels /ɪ, ʊ, ʌ/ were produced as the corner (peripheral) vowels. This indicates that the child was utilizing his L1 Mandarin vowel space as the established base of articulation in building the new L2 vowel system. It is also evident that the child clustered acoustically similar English and Mandarin vowels into several large groups. There were three such vowel clusters: Mandarin /i, y/ and English /i, ɪ, e/ formed the relative high front group; Mandarin /u/ and English /ʊ, u/ formed the high back group; Mandarin /a/ and English /ʌ, ɛ, æ/ formed the relative low-front vowel group. The clustering of several English vowels in the vicinity of Mandarin
corner vowels indicates assimilation of selected L2 categories to established L1 categories and the child’s inability to produce contrasts among them.

The second stage (epochs 2 – 6) represents the Reorganization Phase. In epoch 2, the child started to separate his English vowels from the Mandarin base and this separation suggests a relatively independent development of the L2 vowel system. As can be seen, the child’s English vowels were dispersed more centrally than were the Mandarin vowels which resulted in a smaller L2 working vowel space. This transition from the acoustic overlap with L1 in epoch 1 to a reduction of the L2 space in epoch 2 suggests the emergence of the new L2 vowel system. It represents the child’s attempt to restructure the sound system he has established in his L1 and to develop new L2 corners, away from his Mandarin corner vowels. Beginning from epoch 2, he constructed his L2 space on the basis of his restructured English corner vowels /i, u, æ, a/ and his L2 development can be characterized as a progressive enhancement of this reduced English base in the direction of the corners of monolingual English speakers.

During the next several epochs of the Reorganization Phase the child focused on developing contrasts among individual vowels in L2. To do so, he first separated the two high back vowels /u/-/ʊ/ and the two front vowels /ɛ/-/æ/. He then established the English /ʌ/ in the mid-low back position. Meanwhile, the high-front vowel group /i, ɪ, e/ remained clustered although clearly separated from Mandarin /i/. Mastery of the contrasts among the four back vowels /u, ʊ, o, ɔ/ was particularly difficult for the child and the acquisition of the English /u, o/ distinction was not yet complete even at the end of data collection.
Figure 4.2, showing data points redrawn from Figure 4.1, depicts the great positional variation and acoustic instability of the /u, o/ contrast during the Reorganization Phase.

The development of the /a, ɔ/ contrast was comparatively more typical, starting from a complete acoustic overlap and then undergoing a long process of category separation in the “designated” low back area of the English vowel space. We also note that throughout the Reorganization Phase, the child’s productions of most L2 vowels showed greater variability than his L1 vowels, which was reflected in larger standard errors of the formant frequency means.
Figure 4.2. Developmental trajectories for two back L2 English vowel pairs /ʊ, o/ and /a, ɔ/ produced by the child across eight epochs. Data points are redrawn from Figure 1.
In the third Stabilization Phase (epochs 7 and 8), the child produced distinctively the high front vowels /i, ɪ, e/ which had remained clustered until then. The L2 vowel system has mostly stabilized and, except for a few back vowels, there was little positional variation within the L2 categories. To determine whether the child’s L1 and L2 vowels in the final epoch 8 differed from the systems of English and Mandarin monolinguals, we utilized published data from adult males\(^7\) as a reference for a comparison of the bilingual child’s and monolinguals’ vowel productions.

\(^7\) The adult vowel systems were used as the reference in the present study because they represent the norms which the monolingual children finally achieve. The English vowel space was plotted on the basis of vowels produced by young male adults who spoke the variety of American English typical of central Ohio to which the child was exposed. More details about the speakers and their speech samples can be found in Jacewicz, Fox and Salmons (2007). The Mandarin vowel space was plotted on the basis of vowels produced by native adult Mandarin male speakers reported in Lin & Wang (2001, pp.55). The data from both languages were rescaled to be plotted in the same space.
Figure 4.3. Comparison of the child’s Mandarin and English vowel spaces at epoch 8 with the vowel spaces of corresponding monolingual adults. The left panel shows the dispersion of five monophthongal Mandarin vowels produced by monolingual Mandarin adult male speakers (data reported in Lin & Wang, 2001) and the dispersion of 11 nominal monophthongs in American English produced by monolingual English adult male speakers from the central Ohio area (data reported in Jacewicz, Fox & Salmons, 2007). In the middle panel, the child’s L2 English vowels are superimposed on those of monolingual English adults. In the right panel, the child’s L1 Mandarin vowels are superimposed on those of monolingual Mandarin adults.
In Figure 4.3, the superimposed vowel spaces of the adults (the left panel) show the Mandarin triangular and English quadrilateral shapes, respectively, which constitute the primary difference between the working vowel spaces in these two languages. Certainly, the final L1 and L2 spaces in the child are comparable with the general shapes of the monolingual adults. For a direct comparison, the child’s L2 English vowels are plotted against the vowels of English adults in the middle panel of Figure 4.3. In this display, we see that the two vowel spaces mostly overlap and the general dispersion of L2 vowels is like in the adults, in spite of an apparent counterclockwise shift of the child’s vowel quadrilateral\(^8\). The notable exceptions are for the L2 positions of /e/ and several back vowels, which have still not conformed to those of native speakers. Clearly, the child has established the acoustical distinction between /u/ and /o/ and evidenced a near merger of low back /ɔ/ and /a/, which is typical of younger speakers in central Ohio (e.g., Clopper, Pisoni & de Jong, 2005; Jacewicz, Fox & Salmons, 2011; Labov, Ash & Boberg, 2006; Thomas, 2001). However, he has still not developed the English category /ʊ/ and had difficulty producing the contrast between /ʌ/ and the low back vowels.

L1 Mandarin vowels produced by the child and by the monolingual Mandarin adults are shown in the right panel of Figure 4.3. The general shapes of the two Mandarin vowel spaces are comparable and the corner vowels /i, u, a/ in the child are like those in

---

\(^8\) The counterclockwise shift of the child’s vowel quadrilateral vowel space maybe related to a recent sound change in the vowel system of central Ohio speakers. - This vowel change identified as the North American Shift (Jacewicz, Fox & Salmons, 2011) is present in the speech of younger speakers (mostly children) and it is plausible that this child acquired these new forms through interactions with his playmates and students at the preschool.
Mandarin adults. However, his /ɤ/ and /y/ do not match exactly the positions of the monolingual speakers. To better understand possible sources of this discrepancy, we inspected the developmental paths of the child’s L1 vowels in Figure 4.1. As is evident, his peripheral corner vowels /i, u, a/ remained relatively stable across the eight epochs while the variation was great for the non-peripheral /y/ and /ɤ/. In particular, there was a general trend of /y/-fronting which culminated in epoch 8, when /y/ was located relatively close to /i/. This trend corresponds to the typical route of acquisition in monolingual Mandarin children, who acquire the /y/ late and the developmental back-to-front path is common (Shi & Wen, 2007). Thus, the delay in the child’s production of /i, y/ in their close proximity as in Mandarin adults seems to be related to his continuing L1 development rather than to an influence of the L2.

A different path was in effect for the second vowel /ɤ/, which was produced as a mid, back-central vowel by Mandarin adults. As evident from Figure 4.1, the child first produced it as a relatively raised variant and then lowered and backed it in the vowel space. According to the literature, the /ɤ/ is acquired relatively early in native Mandarin children, usually by age 3 (Si, 2006; Shi & Wen, 2007). Given that the child was 3;8 at the date of the first recording, we conjecture that he had already acquired the vowel at that time and its subsequent positional changes reflect the influence of his emerging L2 English system, particularly his continuous mastery of contrast among several English back vowels.
4.3.2 The change of acoustic vowel space area

The size of the “basic” vowel space (defined by the area of the vowel space bordered by the point vowels) is a parameter often used to characterize the nature of vowel structure differences across prepubertal development of the vocal tract in children as well as across ages, genders, different languages and different dialects of speakers (e.g., Chung et al., 2012; Fox & Jacewicz, 2010; Vorperian & Kent, 2007). We use this measure here to observe the shapes and sizes of the L1 and L2 vowel spaces as a function of L2 exposure. Following a commonly utilized approach, the midpoint formant values of the four corner vowels /i, æ, ɑ, u/ define the vowel space quadrilateral in English (Vorperian & Kent, 2007) and the three corner vowels /i, ɑ, u/ define the vowel space triangle in Mandarin (Chung et al., 2012).

Based on the reports in literature, it is natural that the vowel space area decreases as a function of the developmental increase in vocal tract length (e.g., Vorperian & Kent, 2007; Chung et al., 2012). In order to factor out the effect of vocal tract lengthening, a set of normalized formant frequency values were generated to calculate a normalized vowel space area. We used Lobanov’s (1971) procedure, which converts formant values in Hz to z-scores for each individual speaker. This is a normalization procedure that Adank, Smits, and van Hout (2004) found to be one of the most effective. Since the normalized formant frequency values do not directly reflect Hz values, they were then rescaled into Hz-like values using the method suggested by Thomas and Kendall (2007) to facilitate easier interpretation.
Figure 4.4. The child’s L1 Mandarin triangular vowel space (left) and L2 English quadrilateral vowel space (right) over eight epochs using rescaled normalized formant frequency values.
The change in the Mandarin and English vowel space areas calculated from mean rescaled normalized formant values is plotted in Figure 4.4. The child’s Mandarin vowel space appeared relatively stable across these 20 months. However, his English vowel space showed substantial changes in both the size and general shape—especially during the Reorganization Phase. As his experience with English increased, he continuously modified his pronunciation of English and tended to stabilize his English vowel space.

Figure 4.5. Regression model of the child’s rescaled normalized vowel space areas in L1 Mandarin (filled circles) and L2 English (unfilled circles). The unfilled triangle in epoch 1 is added to the plot to represent the early “English” vowel space defined by the three vowels /i, u, η/.

Shown in Figure 4.5 are scatter plots of the rescaled normalized areas for both Mandarin and English. Superimposed on both plots are regression lines. Since in epoch 1
the child’s English vowel space closely resembled that of the Mandarin triangle, we have plotted two areas for his English space: (1) the triangular area based on the “corner” vowels /i, u, ʌ/ (shown with an open triangle symbol) and (2) the quadrilateral area for consistency with epochs 2 to 8. As can be seen, this English triangular area value is very close to that of the Mandarin triangle, which provides additional support for the claim that in epoch 1, the child was initially basing his L2 vowels on the L1 frame. Regression analysis indicated that there was a significant decline in the Mandarin vowel space across epochs 1 to 8 (F(1,6) = 11.8, p = .014), but that the change was very gradual (-.004 kHz² per epoch). On the other hand, the increase in the size of the English quadrilateral vowel space across epochs 2 to 8 was not only significant (F(1,5) = 27.9, p = .003) but the rate of change was more than 6 times as great in absolute magnitude (.025 kHz² per epoch).

The mostly overlapping Mandarin and English vowel space in epoch 1 and abruptly separated vowel spaces in epoch 2 indicate that the child reduced his English vowel space to maximize the contrast between the languages. This reduction in L2 space was supplemented by a slight expansion of his L1 Mandarin space, most likely to further augment the contrast. As Figure 5 illustrates, there was a steady and fast growth in size of his L2 vowel space after this abrupt reduction in epoch 2 until the final epoch 8, in which the L2 English space area even exceeded his L1 space. On the other hand, the size of his L1 Mandarin space remained relatively unchanged throughout although there was a general slight decline. We can speculate that this declining trend reflected the child’s more casual productions of his Mandarin vowels as a function of L1 experience relative to his “hyperarticulated” L1 vowels at the beginning of the observational period.
4.4 Discussion

This longitudinal case study documented the emergence of bilingualism in a vowel system of a preschool-age boy. Through instrumental analysis of his vowel productions in both L1 and L2, three aspects of his phonetic development were examined: (1) the initial state of his L2 vowel space, (2) the process of L1-L2 separation over the course of his increased exposure to L2, and (3) the status of his L1 vowel system as a function of phonetic category formation in L2.

4.4.1 The initial state of L2 vowel space in an emergent bilingual

The results provided compelling evidence that the 3;8 y.o. child utilized his L1 vowel space as the initial base in building his new L2 vowel system. As expected, the child initially clustered acoustically similar English and Mandarin vowels into several large groups in the vicinity of the corners of his common triangular vowel space. It was striking that the English short non-peripheral vowels (and not the long peripheral vowels) were produced as the upper and lower corners of his “English” triangular vowel space, which suggests that the child relied primarily on similarity in vowel quality (and not quantity) of English vowels relative to his Mandarin vowels as a primary cue in building his English vowel system. At this initial stage, his Mandarin-like triangular “English” vowel space was distinctive from the typical English quadrilateral found in native English-speaking children as young as 1 year of age (Voperian & Kent, 2007). This important finding contributes to the debate about the initial state of L2 in sequential bilinguals, supporting the position that the L1 system initially dominates the L2 (Fantini, 1985; Watson, 1991).
4.4.2 The process of L1-L2 separation

Restructuring of the acoustic vowel space with the emergence of an L2 vowel system can be extensive, especially when phonemic contrasts are to be developed in a crowded English vowel space relative to a sparsely spaced Mandarin system. We expected that the child would initially cluster “similar” L1 and L2 vowel categories in a common acoustic space and that this acoustic overlap would gradually decrease with L1-L2 category separation as a function of his learning phonetic distinctions in L2. Contrary to our prediction, the child diverged from this predicted path very early in his L2 development. In epoch 2, we observed that he drastically reduced the overlap with L1 vowel space and separated the two vowel systems in an unexpected way, by temporarily forming a new periphery in L2 away from the L1 corners. In this way, his new English vowel space became reduced and centralized relative to his established Mandarin corner vowels. This abrupt acoustic shift indicates his attempts to produce the L1-L2 distinctions, primarily for the two high corner vowels /i, u/, and using them as new corners of his emerging L2 space. Accordingly, the phonetic restructuring in epoch 2 illustrates the separation of L1 and L2 vowel spaces, which came about by maximizing the phonetic contrast between L1 and L2. To produce the maximal contrast, the child reduced and centralized the L2 vowel space relative to L1 and even slightly expanded his L1 space by “hyperarticulating” the Mandarin corner vowels. This is an important new result from this study which increases our understanding of strategies used by emergent bilinguals to depart from the established articulatory base in L1.

Throughout the next 16 months, we observed a gradual growth of the reduced L2 vowel space as the child “added” L2 vowel categories to his L2 system by producing a
greater number of acoustic distinctions. At the end of the observational period, the L2 vowel space was not only expanded (and its area was even larger than in L1) but there was basically no acoustic overlap among individual L2 vowels even though not all L2 vowels have yet conformed to the linguistic norms of his American English speech community.

4.4.3 The status of L1 vowel system as a function of phonetic category formation in L2

A particular strength of this study is the use of a longitudinal design to document both the L2 and L1 productions of the same individual over an extended period of time. We assume that the child had remained monolingual until his immersion in English at preschool. Although he was born in the United States and was thus exposed to various sources of auditory sensory information in English, it is unlikely that these sources could have supplied appropriate input for L2 phonetic learning. Kuhl, Tsao and Liu (2003) provided convincing experimental evidence that L2 phonetic learning in infants is not simply triggered by hearing a foreign language. Crucially, early learning is facilitated by social interaction and social contact with a live person who provides information that is referential in nature. It is therefore unlikely that bilingual development of the child began prior to his active exposure to English in the preschool. However, given his young age, we expected his L1 vowel system to be still “flexible” such that it could be restructured under the influence of L2. We also admitted a second possibility, namely that his L1 development was complete and that no further restructuring would take place. The results suggest a third possibility. In particular, while the general shape of his Mandarin vowel space did not change over the observational period and his L1 corner vowels remained
stable throughout, there were considerable ongoing changes in the production of the two non-peripheral vowels /y/ and /ɤ/. Examination of the positional changes of these two vowels led us to propose that the positional variation of /ɤ/ was indeed affected by category formation in L2 but the vowel /y/ was still developing as an L1 phonetic category, suggesting that developmental processes were still active in L1. Thus, the influence of L2 on L1 was limited to one vowel category which overlapped acoustically with several L2 vowels whereas there was no such acoustic interference for the second vowel /y/.

Considering the higher social value of English over Mandarin (according to the child’s mother’s report that he tended to use English as his primary language even at home environment at the end of the recording period), we expect that as the child gains more experience in English, his Mandarin vowels may experience more changes. Not only the non-peripheral vowels, but also the corners vowels may be changed. This process reflects the phenomenon of “subtractive bilingualism” which refers to a gradual loss of the first language due to the acquisition of a more dominant and prestigious second language (Lambert, 1977, 1981; Wong Fillmore, 1991).

4.5 Conclusions

The developmental trajectory of the current sequential bilingual child helps us recognize the complexity of the phonetic restructuring of the vowel system. As a whole, the study contributes the finding that young sequential bilinguals can restructure and separate their two language systems very early although each developmental path may differ. While capturing the emergence of bilingualism in a single child, this study shows
that sequential bilinguals first utilize their existing L1 vowel space and then build their new L2 vowel system “from start,” by means of a drastic restructuring of their existing vowel space to create maximal contrast between the two vowel systems. After this abrupt partitioning, the reduced L2 vowel space is gradually expanded as the L2 learner discovers phonological contrasts in L2 and progresses toward realization of acoustic goals set by native monolingual speakers of that language. Eventually, a sequential bilingual separates the two vowel systems, which suggests separation of the two languages in a bilingual mind.

Although informative, the present findings come from a single child and need to be verified with a greater number of participants. It is also possible that this particular developmental profile is limited to the interaction of phonetic vowel features of Mandarin and English and a different trajectory may result from a contact of languages which both have crowded vowel spaces. Future studies with other bilingual populations are needed to provide further insights into the developmental profile of emergent bilinguals and uncover other possible strategies bilingual children might use to construct their L2 on the basis of L1.
References


Addendum

Word list used to collect Mandarin productions (the target vowels are marked in bold).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Pinyin</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>bi3</td>
<td>pen</td>
</tr>
<tr>
<td></td>
<td>ji1 dan</td>
<td>egg</td>
</tr>
<tr>
<td></td>
<td>yi1</td>
<td>one</td>
</tr>
<tr>
<td></td>
<td>yi1 fu</td>
<td>clothes</td>
</tr>
<tr>
<td>a</td>
<td>ba1</td>
<td>eight</td>
</tr>
<tr>
<td></td>
<td>ma3</td>
<td>horse</td>
</tr>
<tr>
<td></td>
<td>sha fa1</td>
<td>sofa</td>
</tr>
<tr>
<td>o</td>
<td>bo1 cai</td>
<td>spinach</td>
</tr>
<tr>
<td></td>
<td>bo1 luo</td>
<td>pineapple</td>
</tr>
<tr>
<td></td>
<td>bo2 zi</td>
<td>neck</td>
</tr>
<tr>
<td></td>
<td>mo2 gu</td>
<td>mushroom</td>
</tr>
<tr>
<td>y</td>
<td>ge1 zi</td>
<td>pigeon</td>
</tr>
<tr>
<td></td>
<td>she2</td>
<td>snake</td>
</tr>
<tr>
<td></td>
<td>qi che1</td>
<td>car</td>
</tr>
<tr>
<td></td>
<td>dan che1</td>
<td>bike</td>
</tr>
<tr>
<td>u</td>
<td>lao hu3</td>
<td>tiger</td>
</tr>
<tr>
<td></td>
<td>pu2 tao</td>
<td>grape</td>
</tr>
<tr>
<td></td>
<td>shu1 bao</td>
<td>bag</td>
</tr>
<tr>
<td></td>
<td>wu1 gui</td>
<td>tortoise</td>
</tr>
<tr>
<td></td>
<td>zhu1</td>
<td>pig</td>
</tr>
<tr>
<td>y</td>
<td>jin yu2</td>
<td>tropical fish</td>
</tr>
<tr>
<td></td>
<td>lü4 se</td>
<td>green</td>
</tr>
<tr>
<td></td>
<td>yu2</td>
<td>fish</td>
</tr>
</tbody>
</table>
Word list used to collect English productions.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>green</td>
</tr>
<tr>
<td></td>
<td>sheep</td>
</tr>
<tr>
<td></td>
<td>key</td>
</tr>
<tr>
<td>i</td>
<td>pig</td>
</tr>
<tr>
<td></td>
<td>fish</td>
</tr>
<tr>
<td></td>
<td>big</td>
</tr>
<tr>
<td>e</td>
<td>cake</td>
</tr>
<tr>
<td></td>
<td>train</td>
</tr>
<tr>
<td></td>
<td>table</td>
</tr>
<tr>
<td>e</td>
<td>red</td>
</tr>
<tr>
<td></td>
<td>pen</td>
</tr>
<tr>
<td></td>
<td>bed</td>
</tr>
<tr>
<td>æ</td>
<td>cat</td>
</tr>
<tr>
<td></td>
<td>hat</td>
</tr>
<tr>
<td></td>
<td>apple</td>
</tr>
<tr>
<td>u</td>
<td>shoes</td>
</tr>
<tr>
<td></td>
<td>boot</td>
</tr>
<tr>
<td></td>
<td>spoon</td>
</tr>
<tr>
<td>u</td>
<td>book</td>
</tr>
<tr>
<td></td>
<td>foot</td>
</tr>
<tr>
<td></td>
<td>push</td>
</tr>
<tr>
<td>o</td>
<td>nose</td>
</tr>
<tr>
<td></td>
<td>coat</td>
</tr>
<tr>
<td></td>
<td>snow</td>
</tr>
<tr>
<td>ɔ</td>
<td>dog</td>
</tr>
<tr>
<td></td>
<td>doll</td>
</tr>
<tr>
<td>a</td>
<td>watch</td>
</tr>
<tr>
<td></td>
<td>rock</td>
</tr>
<tr>
<td></td>
<td>Stop</td>
</tr>
<tr>
<td></td>
<td>box</td>
</tr>
<tr>
<td>æ</td>
<td>sun</td>
</tr>
<tr>
<td></td>
<td>cup</td>
</tr>
<tr>
<td></td>
<td>truck</td>
</tr>
</tbody>
</table>
Chapter 5: Acoustic Properties of Vowel Production in Bilingual Mandarin-English Children and Age-matched Monolingual Children

Abstract

This study compares the acoustic properties of vowels in bilingual Mandarin-English children with age-matched English and Mandarin monolingual children. Of particular interest how the language exposure to English affects the acoustic features of L1 and L2 vowels in relatively young bilingual Mandarin-English children. Participants included fifteen 5-6 years old bilingual Mandarin-English children (divided into two subgroups based on their proficiency in English), fifteen monolingual Mandarin children and nine monolingual English children. Speech materials included one set of Mandarin word containing /a, i, u, y, ɤ/ and one set of English words containing /i,ɪ, e, ɛ, æ, u,ʊ, o, ɑ, ʌ/. Formant frequencies at five temporal locations (20-35-50-65-80% over the course of vowel duration) were measured using the spectrographic analysis program TF32. Both static (vowel duration, midpoint formant frequency values) and dynamic (formant movement pattern, trajectory length) acoustic properties of the vowels were compared. Bilingual children with limited exposure to English (Bi-low) maintained all acoustic features in their L1 and transferred the L1 features (both static and dynamic) to their L2, which indicates a strong effect of L1 on L2, similar to beginning adulthood L2 learners. Bilingual children with greater exposure to English (Bi-high) produced English vowels in
a near-native manner in both static and dynamic features. However, their production of L1 vowels showed observable positional variation and different patterns of formant movement from monolingual Mandarin children on both similar and dissimilar vowels. In particular, they tended to transfer the features in their L2 vowels to their L1 and moved the L1 vowels closer to L2 vowel, which suggested an assimilatory process associated with the effect of L2 on L1.
5.1 Introduction

The purpose of this study is to examine the extent to which the acoustic properties
of vowels in young bilingual Mandarin-English children are similar to or different from
those of age-matched monolingual children in each language. Of particular interest is
how the interaction between L1 and L2 in young bilingual children affects the phonetic
structures in each language. To date, a large body of research has revealed that early L2
speakers who started to acquire the L2 at a younger age can better produce and
differentiate L2 sounds than late L2 speakers who started to acquire the L2 in adulthood.
(e.g., Aoyama, Guion, Flege, Yamada, & Akahane-Yamada, 2008; Baker, Trofimovich,
Flege, Mack & Halter, 2008; MacLeod, Stoel-Gammon, & Wassink, 2009; MacLeod &
Stoel-Gammon, 2010; Meador, Flege, & MacKay, 2000) In addition, early L2 speakers
are more likely to separate the two phonetic systems compared to late L2 speakers (e.g.,
Guion, 2003). While the majority of these studies examined speech capacity of adult
speakers or adolescents, very few have documented the acoustic profiles of relatively
young bilingual children during the on-going process of language development in each
language. In addition, while many studies have focused on selected vowels; relatively
little has been done to compare the complete vowel inventories of both languages in
bilingual children. The present study will compare the detailed acoustic properties of
vowel production from a relatively complete vowel inventory in bilingual Mandarin-
English children between 5-6 years of age with those of age-matched monolingual
children to examine the effect of the L1-L2 interaction on the similarity/ differences of
young bilingual children’s vowel system relative to the age-matched monolinguals.
Based on the common belief of “earlier is better,” a great number of studies have investigated the extent to which early bilingual speakers can produce or perceive the L2 sounds in a native-like manner. Some researchers have reported that bilinguals who started to acquire an L2 before puberty showed little or no difference in their production of the L2 sounds and their discrimination of L2 contrasts from those of the native speakers (Flege, MacKay & Meador, 1999; Mack, 1989; MacLeod & Stoel-Gammon, 2005). In addition to the monolingual-like ability in L2, some researchers have found that early bilingual speakers were able to retain monolingual ability in their L1. In MacLeod, Stoel-Gammon and Wassink (2009), the formant frequency values of two pairs of high vowels /i-/ and /u-/ produced by bilingual Canadian English-Canadian French speakers were compared with those of corresponding monolingual speakers. The results showed no difference in both F1 and F2 values between bilinguals and each group of monolinguals.

However, some studies have shown that early bilinguals (both simultaneous and sequential bilinguals) may still differ from monolinguals in certain aspects of their production and perception ability (Hojen & Flege, 2006; Pallier, Bosch & Sebastian-Galles, 1997; Sebastian-Galles & Soto-Faraco, 1999; Tsukada, Birdsong, Bialystok, Mack, Sung & Flege, 2005). For example, Sundra, Polka and Baum (2006) compared the voiced-voiceless stop contrasts in early bilingual Canadian English-Canadian French speakers with the corresponding monolingual speakers in terms of the acoustic parameters including VOT, relative burst intensity, mean burst frequency, standard deviation of burst frequency, skewness of burst frequency and kurtosis of burst frequency. The results showed that although bilinguals can produce language-specific differences in
Canadian English and Canadian French stops consonants, these differences were not as large as those produced by respective monolinguals. Additionally, the acoustic parameters of bilingual speakers were similar but not identical to monolinguals.

For those studies which found no differences between early bilinguals and monolinguals of each language, the underlying assumption is that the early bilinguals can develop two separate phonetic systems independently. However, many previous studies have suggested the interdependence of two sub-systems in bilingual speakers (Grosjean, 1989; Keshavarz & Ingram, 2002; Paradis & Genesee, 1996; Paradis, 2001). According to Speech Learning Model (Flege, 1995), a bilingual’s L1 and L2 phonological systems coexist in a common phonological space. Under such a circumstance, L1 and L2 would naturally influence each other. When the interaction of L1 and L2 is strong enough, the bilinguals’ phonetic structures and perceptual organization of both L1 and L2 may change and show different patterns from those of monolinguals of each language.

Generally, there exist two types of interaction effects: the effect of L1 on L2 and vice versa. In terms of the effect of L1 on L2, as Watson (1991) pointed out, for sequential bilinguals who learn a L2 after complete or relatively complete acquisition of L1, their native language (L1) is used as the base to establish the L2 phonetic system. In this case, bilinguals are likely to assimilate the phonetically similar L2 sounds into established L1 categories at the beginning of L2 acquisition. However, not every single L2 sound will be equally assimilated to a L1 sound category. The extent to which the L2 sound is assimilated to the L1 sound category is primarily determined by the phonetic-acoustic similarity between the L2 and L1 sounds (Speech Learning Model, Flege, 1995).
Following continuous exposure and immersion in L2, the influence of L1 on L2 is attenuated and separate L2 sound categories are eventually established.

Unlike the convergence of findings regarding the effect of L1 on L2 in early stage L2 speakers, researchers hold divergent opinions on the influence of L2 on L1. Some studies have reported a change in L1 as a result of L2 learning in both perception (Caramazza, Yeni-Komishian, Zurif & Carbone, 1973; Hazan & Boulakia, 1993) and production (Flege & Hillenbrand, 1984; Flege, 1987; Harada, 2003; Kang & Guion, 2006; Mack, 1990; Williams, 1977). However, other studies, for example, Bohn and Flege (1992) have observed no change of L1 sounds as a function of L2 experience. This study examined the production of vowel /i, ɪ, ɛ/ (similar vowels which occur in both German and English) and /æ/ (which occurs only in English), the production of native German vowels by experienced German speakers of English was not influenced by the long-term immersion in English (L2).

Second, among those studies which reported an influence of L2 on L1, two opposite processes have been observed. One type of mechanism is a dissimilatory process and the other type is an assimilatory process. In terms of the dissimilarity process, it happens when speakers shift an phonetically similar L1 sound away from the L2 sound to make a contrast between the two sounds, establishing a new phonetic category in L2 (Flege, 1995; Flege, Schirru & MacKay, 2003). For example, Guion (2003) found that Quichua (L1) vowels systematically raised and moved away from the similar L2 vowels in Quichua-Spanish bilinguals who have developed distinct vowel categories for the L2 (Spanish). However, researchers have also found instances of a shift in the L1 sound
towards an acoustically similar L2 sound, which represents an assimilatory process. Flege (1987) found the forward movement of French (L1) /u/ as a result of influence from English (L2) /u/ which is located in a more fronted position relative to French /u/. Chang (2012) found that only six weeks’ intensive immersion in Korean (L2) resulted in noticeable assimilatory modification of most English (L1) sounds in native English speakers.

While the underlying driving forces of dissimilatory and assimilatory process of L1 sounds as a function of L2 immersion remains unknown, researchers have shown that the magnitude and direction of L1-L2 interaction in bilinguals were highly correlated with the starting age of L2 learning and the amount of L2 experience. Baker and Trofimovich (2005) compared vowels produced by bilinguals differing in age of exposure to L2 (early and late) and in the amount of experience with L2 (+1 and +7 years) with vowels produced by monolingual speakers. They found that late bilinguals’ L1 (Korean) vowels showed no differences from monolingual Korean speakers while their L2 (English) vowels were quite different from those produced by monolingual English speakers as a result of the influence of their L1. This result indicated that late bilinguals showed a unidirectional influence of their L1 on the L2. However, for the early bilinguals, the inexperienced speakers maintained the monolingual ability in their L1 (Korean) vowels but they were less likely to separate the L2 (English) vowels. The experienced speakers can better separate the L2 (English) vowels but they still differed from monolingual English speakers in certain English vowels. In addition, the experienced speakers also differed from Korean monolinguals on certain L1 (Korean) vowels. These
findings demonstrated that early bilinguals—especially experienced early bilinguals—showed different features from monolinguals of each language as a result of bi-directional influence between their L1 and L2.

A similar bi-directional effect has also been reported in Oh, Guion-Anderson, Aoyama, Flege, Akahane-Yamada and Yamada (2011). This longitudinal study investigated vowel production during English learning by native Japanese adults and children (10 years old) over a one-year period. The results indicated that while Japanese adults showed little change in their production of English vowels, children improved their production of the new English vowels and produced them in a more native-like manner. Meanwhile, the children’s production of Japanese /i/, /a/ and /u/ changed following the mastery of English /i/, /ɪ/, /ɑ/, /ʌ/, /ʊ/.

These studies, along with other research comparing the phonetic features of speech sounds in bilingual speakers with those of monolingual speakers suggest an effect of both age of learning and L2 experience on speakers’ phonetic structure in both languages. However, these findings, to a large extent were based on adults and relatively older childhood L2 speakers. Previous studies have suggested continuing refinement of children’s acoustic properties in their vowel production relative to adults’ norms (Yang & Fox, 2013). Given the on-going process of acoustic development in young children, a more active interaction between L1 and L2 may occur in young bilingual children. In this case, to what extent will young bilingual children’s L2 sound be affected by their L1? To what extent and in which direction will their L1 sounds shift as a result of acquisition of L2 sounds? Will this effect of L2 to L1 influence both similar and dissimilar L1 vowels?
Answering these questions will provide new insight into the process of speech development in bilingual children. To date, cross-language studies in bilingual speech production have been conducted in English-Spanish, English-Korean, English-Japanese and English-French etc. Only a few of them have targeted Mandarin-English bilingual adults (e.g. Chen, Robb, Gilbert & Learman, 2001; Chen, 2006; Jiang, 2008; Wang, 1997). Even less research has investigated the acoustic features in Mandarin-English bilingual children.

The present study aims to examine the effect of L1-L2 interaction on young bilingual Mandarin-English children’s acoustic properties of vowel production relative to age-matched monolinguals. To provide a more detailed comparison of vowel acoustic features, unlike traditional studies that focus just on the midpoint formant frequency values over the course of vowel duration, the present study will also compare the dynamic vowel features in bilingual children with monolingual children or adults. As revealed by some classical work (e.g., Harrington & Cassidy, 1994; Hillenbrand, Getty, Clark and Wheeler, 1995; Neary & Assmann, 1986), vowels show systematic patterns of spectral change regardless of context condition, which characterize the inherent feature of vowel quality. A recent set of studies has demonstrated the importance of vowel spectral change in facilitating vowel perception in adult speakers (Hillenbrand, 2013; Morrison, 2013; Neary, 2013; Strange & Jenkins, 2013). Only a few studies examined the vowel spectral change in native English children (Assmann and Katz; 2000; Assmann, Nearey and Bharadwaj, 2013; Jacewicz, Fox & Salmons, 2011a). To our knowledge, the vowel spectral change in bilingual children has rarely been addressed yet. Therefore, the present
study seeks to add to the body of knowledge of speech production in bilingualism by thoroughly examining both static and dynamic features of vowels in both languages.

Two groups of early bilingual Mandarin-English children (5-6 years) participated in this study. Both groups of children acquired Mandarin earlier than English and thus are sequential bilinguals. One group was highly proficient in English while the other group was less proficient in English. According to above discussion, we hypothesize that the less proficient group of bilingual children will exhibit a strong effect of L1 on the acoustic properties of their L2 vowels. Therefore their L2 will be different from the monolingual children of L2. Regarding the effect of L2 on L1 in less proficient bilingual children, given the active interaction between L1 and L2 in children compared to adults, we expect even the less proficient bilingual children will show some effect of L2 on L1. However, to what extent these less proficient bilingual children maintain their monolingual ability in L1 has yet to be determined. For the group with high proficiency in English, the effect of L1 on L2 is expected to be attenuated. Taking into account of their young age, we expect the high proficient bilingual children will develop monolingual-like ability in producing English vowel features. In addition, the long-term immersion in L2 will be expected to influence their L1. However, due to the lack of evidence in predicting the specific movement direction of the L1 sound under the influence of L2, to what extent and which direction their L1 vowels move are yet to be determined.
5.2 Methods

5.2.1 Speakers

The speakers included 15 sequential Mandarin-English bilingual children (9 girls and 6 boys) aged 5 to 6 years old, 15 age-matched Mandarin monolingual children and 9 age-matched English monolingual children. The bilingual Mandarin-English children were divided into two groups based on their proficiency in English (see Table 5.1 for detailed information). The Bi-high group included 8 children (6 girls and 2 boys) born and raised in the United States (central Ohio regions). These children were raised in a near-monolingual environment and learned Mandarin as the first language from the family contact and interaction with individuals in the local Mandarin-speaking community. Meanwhile, these children had very limited amount of exposure to English before they went to English daycare or kindergarten at about 3 years of age. By the time of recording, these children had had extensive experience with English for about 3 years. The Bi-low group included 7 children (4 girls and 3 boys) born and raised in China who had lived in the U.S. (central Ohio region) for less than 6 months. Compared to Bi-high group of children, the Bi-low group of children started to learn English at a slightly older age and had less exposure in English. In terms of dialect background of bilingual children’s parents, 13 out of 15 children have both parents coming from northern dialect regions of China. The other two children have one parent from northern dialect regions. Mandarin is the daily-life language used in all bilingual children’s family. The 15 age-matched monolingual Mandarin speakers were native Mandarin speakers born and raised in the Beijing area. The 9 monolingual English children were native English speakers
born and raised in central Ohio region. All children were reported as having no speech and language disorders.

<table>
<thead>
<tr>
<th>Group</th>
<th>SubNum</th>
<th>Gender</th>
<th>Age</th>
<th>LOR_in_US</th>
<th>AOL_E</th>
<th>E_usage</th>
<th>AOL_M</th>
<th>M_usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-high</td>
<td>107</td>
<td>F</td>
<td>5;11</td>
<td>58m</td>
<td>3;6</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>108</td>
<td>F</td>
<td>5;7</td>
<td>67m</td>
<td>4</td>
<td>50%</td>
<td>since born</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>118</td>
<td>M</td>
<td>5;5</td>
<td>65m</td>
<td>3;7</td>
<td>80%</td>
<td>since born</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>119</td>
<td>F</td>
<td>5;10</td>
<td>70m</td>
<td>2;1</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>123</td>
<td>F</td>
<td>5;10</td>
<td>47m</td>
<td>3</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>124</td>
<td>F</td>
<td>5;6</td>
<td>66m</td>
<td>2;6</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>125</td>
<td>M</td>
<td>5;1</td>
<td>45m</td>
<td>2</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>F</td>
<td>5;7</td>
<td>66m</td>
<td>3</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>131</td>
<td>F</td>
<td>5;7</td>
<td>66m</td>
<td>3</td>
<td>70%</td>
<td>since born</td>
<td>30%</td>
</tr>
<tr>
<td>Bi-low</td>
<td>109</td>
<td>F</td>
<td>5;10</td>
<td>5m</td>
<td>4;6 (in China), 5;5 (in US)</td>
<td>40%</td>
<td>since born</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>M</td>
<td>5;8</td>
<td>3m</td>
<td>4 (in China), 5;5 (in US)</td>
<td>40%</td>
<td>since born</td>
<td>60%</td>
</tr>
<tr>
<td></td>
<td>111</td>
<td>M</td>
<td>5;8</td>
<td>3m</td>
<td>5;5 (in US)</td>
<td>30%</td>
<td>since born</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>M</td>
<td>5;2</td>
<td>1.5m</td>
<td>4 (in China), 5 (in US)</td>
<td>30%</td>
<td>since born</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>F</td>
<td>5;6</td>
<td>6m</td>
<td>3 (in China), 5 (in US)</td>
<td>20%</td>
<td>since born</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>126</td>
<td>F</td>
<td>6;3</td>
<td>5m</td>
<td>5;11 (in US)</td>
<td>20%</td>
<td>since born</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>127</td>
<td>F</td>
<td>5;10</td>
<td>4m</td>
<td>5;6 (in US)</td>
<td>40%</td>
<td>since born</td>
<td>60%</td>
</tr>
</tbody>
</table>

Table 5.1. Characteristics of subgroups of bilingual Mandarin-English children: subject group, subject number, gender, age, length of residency in US, age of learning in English, English usage, age of learning in Mandarin, Mandarin usage.

5.2.2 Speech materials

The recording materials for bilingual children consisted of two word lists: 10 Mandarin disyllabic words containing 5 Mandarin monophthongal vowel phonemes /a, i, u, y, ɤ/ and 20 English monosyllabic/disyllabic words containing 10 American English
vowels /i, ɪ, e, ɛ, æ, u, ʊ, o, ɑ, ʌ/. In all disyllabic words, the target vowel appeared in the first syllable. The selection of both Mandarin and English words was based on word familiarity, word frequency (Thorndike & Lorge, 1963) and picturability. The phonetic context of both Mandarin and English words were controlled. In addition, the third tone was excluded for Mandarin words to avoid the longer duration of vowels in syllables with tone 3.

5.2.3 Procedures

There were two recording sessions for each bilingual subject. Mandarin words were produced in the first session and English words were produced in the second session after a 15-20 minute break. The experimenter (a fluent bilingual Mandarin-English speaker) interacted with the speakers in Mandarin during Mandarin session and English during English session. For each monolingual speaker, only one recording session was conducted in their native language. In each recording session, a visual-auditory word repetition task was used to collect speech samples under control of a custom Matlab program. During the recording period, each speaker was seated facing a laptop computer.

---

9 Due to the merger of /ɔ/ and /ɑ/ in most dialects of American English including Ohio English (Clopper & Pisoni, 2004; Clopper, Pisoni & de Jong 2005, Labov, Ash & Boberg 2006), just one vowel /ɑ/ was selected for use in the present study.

10 The consonants preceding English vowels were all stops except for /f/ in “feet” while the following consonants of were all voiceless stops or fricative /s/. As for Mandarin, due to the Mandarin syllable structure rule, /y/ cannot follow stops; therefore, the words containing /y/ has either zero initial (no preceding consonant) or the affricate /ʨ/.

11 Vowels in a syllable with the third tone have significantly longer duration than those in a syllable with other three tones in Mandarin (Chang, 2010; Lin, 1988; Tseng, 1981; Xu, Tsai, & Pfingst, 2002).
in a quiet room. Randomly ordered\textsuperscript{12} pictures containing target words were presented on the computer screen followed by an audio prompt produced by a native adult speaker. The participants were then asked to repeat each word immediately after the prompt. During these sessions, a Shure SM10A head-mounted microphone was situated approximately 1-inch from the subject’s mouth. Speech samples were recorded directly onto a hard drive disk with a 16-bit quantization rate and 44.1 kHz sampling rate.

5.2.4 Acoustic measurements

5.2.4.1 Vowel duration

Measurements of vowel duration were based on the landmark locations of vowel onset and offset located by hand and determined primarily on the basis of the waveform, accompanying with visual check of the spectrogram (Kent & Read, 1992). Vowel onset was defined at the point of zero crossing of the first glottal following stop closure release or cessation of frication. For the Mandarin word “yú” (fish) started with a zero initial, the vowel onset was defined as the start of first clear glottal pulse. Vowel offset for Mandarin vowels was set at the point of the zero crossing point of the last glottal pulse extending through F1 and F2. The offset for English vowels was set at the zero crossing point close to the significant decrease in amplitude due to the stop closure or was marked at the point prior to the onset of frication.

\textsuperscript{12} The same random order was used for both recording blocks.
5.2.4.2 Formant frequencies

All tokens were first down-sampled to 11.25 kHz using a custom Matlab program. The spectrographic analysis program TF32 (Milenkovic, 2003) was then used to extract the frequency values of the first two formants, F1 and F2 at five equidistant temporal locations (20-35-50-65-80% point) in order to capture the dynamic spectral change of the vowel movement (Fox & Jacewicz, 2009; Jacewicz, Fox & Salmons, 2011b). Since the speakers across all three groups were similar in age, the effect of different vocal tract lengths on the formant frequencies was expected to be small. Thus, unnormalized formant frequency values were used for further calculation and analysis.

5.2.4.3 Trajectory length

The first measurements derived on the basis of formant frequency values was trajectory length (TL). TL is the sum of the Euclidean distances (in Hz) of the four separate vowel sections between 20% and 80% point (Fox & Jacewicz, 2009),

\[ TL = \sum_{n=1}^{4} VSL_n \]  

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

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

A longer TL reflects greater change in formant frequency values and greater magnitude of formant movement. This measure provides a detailed evaluation of the extent to which the formant trajectory changes over the course of vowel duration between the selected time locations. This measure is particularly helpful in capturing the formant movement of
the curved formant trajectory which has formant frequency values that return in later temporal portions to those in the earlier temporal portions (these trajectories “curve back”).

5.3 Results

5.3.1 Mandarin vowel system

5.3.1.1 Vowel duration

Mean vowel durations for each Mandarin vowel produced by bilingual Mandarin-English and age-matched monolingual Mandarin children are presented in Figure 5.1 and Table 5.2. In general, both groups of bilingual children produced vowel durations similar to those produced by monolingual Mandarin children. The effect of vowel quality on duration is anticipated but is not of particular interest in the present study. Therefore, one-way ANOVA was used to examine the effect of English proficiency on the duration of each separate Mandarin vowel. The results revealed no significant effect of L2 proficiency on bilingual children’s production of vowel duration in their native language relative to monolingual Mandarin children. This indicates that the language proficiency in L2 does not change bilingual children’s vowel duration in their native language relative to the monolingual children.
Figure 5.1. Bar plots showing mean and standard error of vowel duration for each Mandarin vowel in each group of children.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Bi-low M (SE)</th>
<th>Bi-high M (SE)</th>
<th>Mono M (SE)</th>
<th>p</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>301 (14)</td>
<td>293 (12)</td>
<td>275 (12)</td>
<td>0.236</td>
<td>0.073</td>
</tr>
<tr>
<td>i</td>
<td>282 (15)</td>
<td>268 (12)</td>
<td>299 (10)</td>
<td>0.18</td>
<td>0.119</td>
</tr>
<tr>
<td>u</td>
<td>250 (17)</td>
<td>249 (10)</td>
<td>271 (12)</td>
<td>0.379</td>
<td>0.069</td>
</tr>
<tr>
<td>y</td>
<td>283 (23)</td>
<td>290 (7)</td>
<td>293 (13)</td>
<td>0.903</td>
<td>0.007</td>
</tr>
<tr>
<td>ɤ</td>
<td>295 (16)</td>
<td>314 (13)</td>
<td>330 (11)</td>
<td>0.172</td>
<td>0.122</td>
</tr>
</tbody>
</table>

Table 5.2. Summary of statistical results showing the effect of L2 experience on vowel duration for each Mandarin vowel (means and standard errors of duration are in ms).

5.3.1.2 Midpoint F1 by F2 vowel space

Following the traditional approach of using midpoint formant frequency values to characterize the basic acoustic features of vowel production, scatter plots of the midpoint locations for all vowels were measured and plotted (shown in Figure 5.2) in the F1 x F2
vowel space. To better describe and quantify the distribution of each vowel category, vowel ellipses were plotted for individual vowel categories following the method used in Zhou and Xu (2008). For each vowel, the center of the ellipse was set at the mean of F2 along the fitting line which was determined by the positive angle of the linear fit of the data points. The semimajor axis was at the same direction as the fitting line and the length of it was set at two standard deviations from the ellipse center. The semiminor axis was perpendicular to the fitting line and the length of it was set at two standard deviations from the ellipse center. Each ellipse encircled approximately 95% of the samples in each Mandarin vowel category.

Both groups of bilingual children clearly separated the five Mandarin vowels in the acoustic vowel space. However, Bi-high children showed greater variation in their production of /y/ and some overlap between /i/ and /y/ relative to the Bi-low and Mono children. In addition, Bi-high children also produced /a/ at a higher position than Bi-low and monolingual children. For the vowel /ɤ/, both Bi-low and Bi-high children showed greater variation than Mono children. While for the vowel /u/, Bi-low children showed larger variation than Bi-high and Mono children. The ellipse area of each vowel category is shown in Table 5.3. All three groups of children had similar ellipse size for the vowel /a/ and /i/. While for the vowel /y/ and /ɤ/, Bi-high children had a larger ellipse area than the other two groups, which was consistent with the more scattered distribution of this vowel in the scatter plots. For the vowel /u/, Bi-low children had a larger vowel ellipse than the other two groups which was consistent with a more scattered distribution of this vowel in the scatter plots.
Figure 5.2. Scatter plots of midpoint formant frequency values for individual Mandarin vowel categories in each group of children.
Table 5.3. The ellipse areas of individual Mandarin vowels in each group of children (in kHz²).

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>y</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low</td>
<td>0.213</td>
<td>0.100</td>
<td>0.193</td>
<td>0.155</td>
<td>0.287</td>
</tr>
<tr>
<td>Bi-high</td>
<td>0.250</td>
<td>0.095</td>
<td>0.063</td>
<td>0.312</td>
<td>0.292</td>
</tr>
<tr>
<td>Mono</td>
<td>0.272</td>
<td>0.128</td>
<td>0.076</td>
<td>0.231</td>
<td>0.155</td>
</tr>
</tbody>
</table>

Figure 5.3 shows the mean and standard error of midpoint formant frequency values for each vowel in each group of children. In order to examine the extent to which the language experience affected the relative positions of vowels produced by bilingual children relative to those by monolingual children in the F1 x F2 acoustic space, the Euclidean distance between Bi-low and Mono as well as Bi-high and Mono for each vowel category was calculated. As shown in Table 5.4, the distances between Bi-low and Mono children were much smaller than those between Bi-high and Mono children for all five Mandarin vowels. This indicates that Bi-low children’s production of Mandarin vowels were acoustically more similar to Mono children. Then one-way ANOVA was used to examine whether the three groups were different on each formant (F1 and F2) for each vowel. The statistical results showed a significant effect of language experience on F1 of /y/ (F(2,27) = 3.471, \( p = 0.046, \eta^2 = 0.205 \)) and F2 of /ɤ/ (F(2,27) = 3.988, \( p = 0.03, \eta^2 = 0.228 \)). In particular, Tukey HSD revealed that Bi-high children produced significantly higher F1 for the vowel /y/ and higher F2 for /ɤ/ than Mono children. These results demonstrated that language experience did not change the relative position of native vowels in the acoustic space in Bi-low children. However, it influenced the relative position of certain vowels in Bi-high children.
Figure 5.3. Dispersion of Mandarin vowels for each group of children plotted on the basis of the means and standard errors of midpoint formant frequency values.

Table 5.4. Euclidean distance between the centers of each vowel category of bilingual group with that of monolingual children for each Mandarin vowel (distances are in Hz).

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>i</th>
<th>u</th>
<th>y</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low_Mono</td>
<td>20.30</td>
<td>17.35</td>
<td>42.26</td>
<td>26.78</td>
<td>164.09</td>
</tr>
<tr>
<td>Bi-high_Mono</td>
<td>127.33</td>
<td>84.64</td>
<td>72.66</td>
<td>170.36</td>
<td>176.35</td>
</tr>
</tbody>
</table>

5.3.1.3 Vowel spectral change of Mandarin vowels

While the vowel dispersion pattern of the midpoint formant frequency values represents a traditional approach to describe the general distribution of relative position in the acoustic vowel space, it does not allow us to capture the nature of vowel quality change over time. As mentioned earlier, measurements made at multiple time locations...
(e.g. 20-35-50-65-80% point) provide an estimate of the extent of formant movement through the central 60% portion of vowel duration. Based on the multiple measurements, the direction and the magnitude of the formant movement can be described. As shown in Figure 5.4, both Bi-low and Bi-high children showed similar direction of formant movement for most of Mandarin vowels. However, Bi-high children showed different shapes and/or magnitudes of formant movement than did the Mono and Bi-low children for the vowels /i/, /y/, /a/ and /ɤ/.

**Figure 5.4.** Vowel dynamic 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 five Mandarin vowels in each group of children. The larger size symbols represent 80% point.
More details of the spectral change patterns are presented in Figure 5.5. The vowel /i/ showed a moderate increase in F1 and decrease in F2 in Bi-low and Mono children. However, in Bi-high children this vowel was more fronted and showed smaller changes in F1 and F2. The vowel /y/ showed consistent increase in F1 in Mono and Bi-low children but very small change of F1 in Bi-high children. In addition, this vowel was fronted and became closer to /i/ in Bi-high children than those produced by the other two groups. For the vowel /a/, in all three groups of children, F1 increased from 20-50% point and then decreased from 50-80% point. F2 decreased from 20-50% point and then increased from 50-80% point. However, the formant trajectory of Bi-high children showed a crossover and was located at a higher position with lower F1 than those in the other two groups. For the vowel /ɤ/, in both Bi-low and Mono children, there was a great magnitude increase in F1 coupled with a slight increase in F2. However, in Bi-high children, although F2 showed gradual increase pattern, the F1 increased to a much smaller extent. Therefore, the vowel /ɤ/ showed a less magnitude of formant movement in Bi-high children relative to the other two groups. In general, it can be observed that Bi-low had formant trajectories similar to those of the Mono children. However, vowels produced by Bi-high children differed from Mono children in both relative position and the magnitude of formant movement.
Figure 5.5. Illustration of vowel dynamic spectral change for subsets of Mandarin vowels in each group of children. Note that the scale of each panel is different. The larger size symbols represent 80% point.
Figure 5.6 and Table 5.5 shows the means and standard errors of the trajectory length for each Mandarin vowel for all three groups of children. As can be seen, Bi-low children have trajectory lengths similar to those of the Mono children for all five vowels while Bi-high children produced shorter vowel trajectory lengths than did Bi-low and Mono children. One-way ANOVA and Tukey HSD were used to determine the group differences of the trajectory lengths for individual Mandarin vowels (summarized in Table 5.5). The results show no significant differences of trajectory lengths between Bi-low and Mono children for any vowel. However, for the vowel /i, y, ɤ/, Bi-high children showed significantly shorter trajectory lengths than did the Mono children. These results indicate that language experience in English produced a significant change in the vowel dynamics of Bi-high children in their native vowels. However, the short period of immersion in English did not affect Bi-low children’s vowel dynamic features in their Mandarin vowels.
Figure 5.6. Bar plots showing the mean and standard error of trajectory length for each Mandarin vowel in each group of children.

Table 5.5. Summary of statistical results showing the effect of L2 experience on trajectory length of each Mandarin vowel (means and standard errors are in Hz).
5.3.2 English vowel system

5.3.2.1 Vowel duration

Mean vowel durations of English vowels produced by bilingual and age-matched monolingual English children is presented in Figure 5.7 and Table 5.6. Among these three groups of children, vowels produced by Bi-low children had longer durations than those produced by Bi-high and Mono children for nearly all 10 vowels (the single exception is /a/). To better understand how the bilingual children differ from monolingual English children in the duration of individual English vowels\(^{13}\), one-way ANOVA and Tukey HSD were used to examine the group differences (summarized in Table 5.6). The results showed no significant difference between Bi-high and Mono children. But Bi-low children produced vowel /i/ and /ɪ/ with significantly longer duration than those produced by Bi-high or Mono children.

\(^{13}\) The effect of vowel quality on the duration in English has been demonstrated in previous studies (Jacewicz, Fox & Salmons, 2007) and is not of particular interest in the present study.
Figure 5.7. Bar plots showing mean and standard error of duration for each English vowel in each group of children

Table 5.6. Summary of statistical results showing the effect of L2 experience on vowel duration for each English vowel (means and standard errors are in ms).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Bi-low M (SE)</th>
<th>Bi-high M (SE)</th>
<th>Mono M (SE)</th>
<th>p</th>
<th>η²</th>
<th>Significant difference</th>
<th>Tukey HSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>234 (18)</td>
<td>174 (9)</td>
<td>186 (10)</td>
<td>0.007</td>
<td>0.374</td>
<td>Bi-high&lt;Bi-low, Mono&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>149 (15)</td>
<td>110 (8)</td>
<td>119 (8)</td>
<td>0.031</td>
<td>0.281</td>
<td>Bi-high&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>191 (20)</td>
<td>173 (13)</td>
<td>179 (13)</td>
<td>0.721</td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>æ</td>
<td>178 (18)</td>
<td>155 (7)</td>
<td>147 (9)</td>
<td>0.192</td>
<td>0.146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ð</td>
<td>204 (20)</td>
<td>193 (10)</td>
<td>182 (12)</td>
<td>0.542</td>
<td>0.057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>207 (20)</td>
<td>205 (13)</td>
<td>180 (11)</td>
<td>0.338</td>
<td>0.098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ø</td>
<td>120 (11)</td>
<td>102 (8)</td>
<td>110 (7)</td>
<td>0.354</td>
<td>0.094</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>196 (21)</td>
<td>182 (4)</td>
<td>178 (10)</td>
<td>0.564</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>173 (16)</td>
<td>189 (23)</td>
<td>166 (14)</td>
<td>0.638</td>
<td>0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ʌ</td>
<td>154 (12)</td>
<td>131 (21)</td>
<td>129 (14)</td>
<td>0.523</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.3.2.2 Midpoint F1 by F2 vowel space

Figure 5.8 shows the scatter plots based on F1 and F2 at midpoint location for each English vowel in each group of children. Following the same method used for Mandarin vowels, an ellipse was plotted for each English vowel category. Unlike Bi-high and Mono children, Bi-low children showed significant overlaps of these ellipses in the F1 by F2 vowel space. In general, the production of the ten English vowels in Bi-low children can be categorized into four clusters in the acoustic vowel space: /i, ɪ, e/, /ε, æ/, /u, ŭ, o/ and /ɑ, ʌ/. These highly overlapped vowel clusters indicate that Bi-low children did not clearly separate these English vowels in the acoustic space. As shown in the middle panel, Bi-high children showed better separation among the vowel categories. A similar pattern to that of Bi-high children was also observed in Mono children. Although Bi-high children were similar to Mono children in their better separated acoustic vowel categories, they still demonstrate differences from Mono children. In particular, the vowel /ɪ/ produced by Bi-high children was more scattered than that produced by Mono children. However, vowel /u/ and /û/ produced by Bi-high children are more concentrated than those produced by Mono children.
Figure 5.8. Scatter plots of midpoint formant frequency values for individual English vowel categories in each group of children.
To examine the extent to which the vowel categories were separated, for each group, the proportion of each overlap to each vowel of the selected intersected vowel pairs (i-ɪ, i-e, i-e, u-u, u-o, u-o, e-æ, æ-ʌ) was calculated (e.g. shown in Figure 5.9). First, the ellipse area of each vowel category was calculated (shown in Table 5.7). Then, the overlap area of each selected vowel pairs was calculated (shown in Table 5.8), similar to the method of SOAM 2-D in Wassink (2006). Finally, the proportion of each overlap to each of the two involved vowel categories was calculated (shown in Table 5.9). As can be seen, the areas of overlaps for most vowel pairs in Bi-low group were substantially greater than the other two groups of children. In addition, the proportions of overlap to each vowel category in Bi-low children were also greater than those in the other two groups. The greater overlaps in Bi-low children indicate that they are less likely to separate the vowel categories in the acoustical space.

Figure 5.9. Illustration of overlap between /ɪ/ and /e/ in Bi-low group of children.
<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>i</th>
<th>e</th>
<th>e</th>
<th>æ</th>
<th>u</th>
<th>o</th>
<th>o</th>
<th>a</th>
<th>ʌ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low</td>
<td>0.143</td>
<td>0.309</td>
<td>0.246</td>
<td>0.443</td>
<td>0.399</td>
<td>0.230</td>
<td>0.492</td>
<td>0.220</td>
<td>0.245</td>
<td>0.294</td>
</tr>
<tr>
<td>Bi-high</td>
<td>0.114</td>
<td>0.280</td>
<td>0.150</td>
<td>0.158</td>
<td>0.276</td>
<td>0.258</td>
<td>0.231</td>
<td>0.169</td>
<td>0.157</td>
<td>0.158</td>
</tr>
<tr>
<td>Mono</td>
<td>0.234</td>
<td>0.214</td>
<td>0.299</td>
<td>0.169</td>
<td>0.206</td>
<td>0.355</td>
<td>0.340</td>
<td>0.202</td>
<td>0.139</td>
<td>0.219</td>
</tr>
</tbody>
</table>

Table 5.7. The area of ellipses of each English vowel in each group of children (in kHz²).

<table>
<thead>
<tr>
<th></th>
<th>i-i</th>
<th>i-e</th>
<th>i-e</th>
<th>u-o</th>
<th>u-o</th>
<th>u-o</th>
<th>e-æ</th>
<th>æ-ʌ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low</td>
<td>0.030</td>
<td>0.052</td>
<td>0.190</td>
<td>0.206</td>
<td>0.089</td>
<td>0.220</td>
<td>0.299</td>
<td>0.194</td>
</tr>
<tr>
<td>Bi-high</td>
<td>0.004</td>
<td>0.006</td>
<td>0.082</td>
<td>0.057</td>
<td>0.006</td>
<td>0.114</td>
<td>0.024</td>
<td>0.044</td>
</tr>
<tr>
<td>Mono</td>
<td>0.000</td>
<td>0.095</td>
<td>0.095</td>
<td>0.096</td>
<td>0.035</td>
<td>0.171</td>
<td>0.065</td>
<td>0.071</td>
</tr>
</tbody>
</table>

Table 5.8. The area of overlap for eight selected vowel pairs in each group of children (in kHz²).

<table>
<thead>
<tr>
<th></th>
<th>O/i</th>
<th>O/i</th>
<th>O/e</th>
<th>O/e</th>
<th>O/u</th>
<th>O/u</th>
<th>O/u</th>
<th>O/o</th>
<th>O/o</th>
<th>O/e</th>
<th>O/æ</th>
<th>O/ʌ</th>
<th>O/ʌ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low</td>
<td>21%</td>
<td>10%</td>
<td>37%</td>
<td>21%</td>
<td>61%</td>
<td>77%</td>
<td>89%</td>
<td>42%</td>
<td>39%</td>
<td>40%</td>
<td>45%</td>
<td>100%</td>
<td>68%</td>
</tr>
<tr>
<td>Bi-high</td>
<td>4%</td>
<td>1%</td>
<td>5%</td>
<td>4%</td>
<td>29%</td>
<td>55%</td>
<td>22%</td>
<td>25%</td>
<td>2%</td>
<td>4%</td>
<td>49%</td>
<td>68%</td>
<td>15%</td>
</tr>
<tr>
<td>Mono</td>
<td>0</td>
<td>0</td>
<td>41%</td>
<td>32%</td>
<td>44%</td>
<td>32%</td>
<td>27%</td>
<td>28%</td>
<td>10%</td>
<td>18%</td>
<td>51%</td>
<td>85%</td>
<td>38%</td>
</tr>
</tbody>
</table>

Table 5.9. The proportion of each overlap to each vowel category for all eight vowel pairs in each group of children.

Figure 5.10 shows the means and standard deviations of midpoint formant frequency values for each English vowel in each group of children. Again, the Euclidean distance between Bi-low and Mono as well as Bi-high and Mono for each vowel category was calculated to examine the extent to which the language experience affected the vowel
positions in bilingual children relative to monolingual children in the F1 x F2 acoustic space. As shown in Table 5.10, the distances between Bi-high and Mono children were much smaller than those between Bi-low and Mono children for almost every English vowel besides /o/ and /æ/. This indicates that Bi-high children’s production of English vowels were acoustically more similar to Mono children. Then, one-way ANOVA was conducted on each of the formants (F1 and F2) for each individual vowel. The results showed a significant difference on F2 for the vowel /i/ (F(2,21) = 3.489, p = 0.049, η² = 0.249) and /u/ (F(2,21) = 6.399, p = 0.007, η² = 0.379) across three groups of children. Tukey HSD results showed that Bi-low children produced a lower F2 than Mono children for the vowel /i/. They produced a lower F2 than both Bi-high and Mono children for the vowel /u/. These results indicate that Bi-high children are similar to Mono children in the relative position of vowel categories. However, due to the scattered distribution of vowel categories in Bi-low children, the statistical results did not provide much evidence about their difference from the Mono children even though several sporadic significant results were obtained between Bi-low and Mono children.
Figure 5.10. Dispersion of English vowels for each group of children plotted on the basis of the means and standard errors of rescaled normalized formant frequency values at midpoint location.

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>i</th>
<th>e</th>
<th>e</th>
<th>æ</th>
<th>u</th>
<th>o</th>
<th>o</th>
<th>a</th>
<th>æ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low_Mono</td>
<td>223.60</td>
<td>70.52</td>
<td>301.64</td>
<td>86.14</td>
<td>51.32</td>
<td>474.14</td>
<td>221.12</td>
<td>5.08</td>
<td>78.21</td>
<td>98.97</td>
</tr>
<tr>
<td>Bi-high_Mono</td>
<td>18.60</td>
<td>21.81</td>
<td>60.23</td>
<td>64.95</td>
<td>70.04</td>
<td>64.54</td>
<td>82.55</td>
<td>91.40</td>
<td>52.91</td>
<td>61.15</td>
</tr>
</tbody>
</table>

Table 5.10. Euclidean distance between the mean of each of bilingual group with that of monolingual children for each English vowel (distances are in Hz).
5.3.2.3 *Vowel spectral change of English vowels*

Figure 5.11 shows the formant trajectories of these ten English vowels for each group. Among these three groups of children, Bi-high children showed similar formant movement patterns to Mono children for the majority of the English vowels. However, Bi-low children differed from Bi-high and Mono children for all English vowels in terms of the relative positions, direction and/or magnitudes of the formant trajectories. To better compare the formant trajectories across three groups of children in more details, these ten English vowels were divided into four classes and each of them is shown in Figure 5.12.

Figure 5.11. Vowel dynamic 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 in each group of children. The larger size symbols represent 80% point.
For the three high front vowels /i, ɪ, e/, the formant trajectory of /i/ showed a slight increase in F2 coupled with a slight decrease in F1 in Bi-high and Mono children. However, in Bi-low children, it showed an increase followed by a decrease in F2. Therefore, the direction of formant movement in Bi-low children was very different from that in the other two groups. In addition, the vowel /i/ was located at a relatively back and lower position in B-low children compared to those in Bi-high and Mono children. In the case of /ɪ/, all three groups of children showed a different pattern of formant movement. Mono children showed little change in F2 but an increase followed by a decrease in F1. Bi-high children showed a decrease in F2 but little change in F1. In contrast, Bi-low children showed a decrease in F1 but no change in F2. In the case of /e/, there was a considerable decrease in F1 with an increase in F2 in Bi-high and Mono children. This pattern clearly demonstrated the nature of diphthongization of /e/ in English. However, in Bi-low children, the vowel /e/ showed less formant movement. In addition, /e/ and /ɪ/ in Bi-low children showed greater positional proximity than they did in Bi-high and Mono children.

For the two low front vowels /ɛ/ and /æ/, /e/ in Bi-high and Mono children showed an increase in F1 with little change in F2. However, in Bi-low children, it was located at a relatively back and lower position with a reversal direction in F1 and small decrease in F2 which produced a curved back formant trajectory. /æ/ in Bi-high and Mono children also showed an increase in F1 with slight decrease in F2. However, it
showed a curved-back formant movement with reversal direction of F1 coupled with a slight increase in F2 in Bi-low children. The direction of formant movement of /æ/ in Bi-low children was opposite to those in the other two groups. In addition, these two vowels showed greater positional proximity in Bi-low children than they did in Bi-high and Mono children.

For the high back vowels /u, ʊ, o/, /u/ showed a substantial decrease in F2 with slight decrease in F1 in both Bi-high and Mono children, which demonstrated a central-to-back movement of the formant trajectory. However, in Bi-low children, this vowel showed much less formant movement with no change in F1 and slight increase in F2. In addition, this vowel was located at a back position and the direction of the formant trajectory of this vowel was opposite to those in Mono and Bi-high children. For the vowel /ʊ/, there was little change in either F1 or F2 across all three groups of children. However, this vowel produced by Bi-low children was located at a relatively back and higher position. For the vowel /o/, there was a dramatic decrease in F1 coupled with moderate decrease in F2 in Bi-high and Mono children which characterizes the expected diphthongal nature of this vowel in American English. However, in Bi-low children, it showed little formant movement which indicates a relatively static feature of the vowel quality.

For the low back vowels /ʌ/ and /ɑ/, /ʌ/ showed a slight increase in F1 and some decrease in F2 in Bi-high and Mono children. In contrast, in Bi-low children, this vowel was lowered and showed little change in F1. Similar to /ʌ/, /ɑ/ also showed an increase in
F1 and decrease in F2 in Bi-high and Mono children. However, in Bi-low children, /ɑ/ was fronted and characterized by a curved-back formant movement with a reversal direction of F1 coupled with a decrease in F2. In addition, these two vowels were in closer proximity in Bi-low children compared to those in the other two groups. These positional approximations of English vowels in Bi-low children were consistent with the pattern observed in their vowel scatter plot.

Figure 5.12. Illustration of English vowel spectral change for four subsets of vowels in each group of children. Note that the scale of each panel is different. The larger size symbol represents 80% point.
Figure 5.13 presents the means and standard errors of trajectory length for each English vowel in each group (also summarized in Table 5.11). It can be observed that the trajectory lengths of Bi-high children were similar to those of Mono children for most vowels. For example, both Bi-high and Mono children produced shorter trajectory length for the relatively static vowels (such as /i, ɪ, u/) compared to the diphthongized vowels (such as /e/ and /o/). However, Bi-high children still showed differences from Mono children. In particular, for the vowels /i, ɪ, u, a/, Bi-high children produced longer trajectory lengths than Mono children. Compared to Bi-high and Mono children, Bi-low children produced similar TL for most of the English vowels. One-way ANOVA and Tukey HSD were used to examine whether bilingual children differed significantly from monolingual English children in the trajectory length of each English vowel. The results showed a complex pattern of group differences. In particular, Bi-low children produced significantly longer trajectory lengths for the vowels /i/, /ɪ/, /ɛ/, /æ/, /a/ while shorter trajectory lengths for the vowel /o/ than did Mono children. These differences between Bi-low and Mono children demonstrated that these young children learning English with a short period of time (less than 6 months in the present study) had not developed native-like vowel spectral change patterns as monolingual English children. On the other hand, while Bi-high children showed limited differences from Mono children on most English vowels, they produced significantly longer trajectory lengths for the vowel /ɪ/ than did Mono children. This suggested that even bilingual children who are highly proficient in
English and have used English for a long time still differ from monolingual peers in certain aspect of the acoustic properties.

Figure 5.13. Bar plots showing the mean and standard error of trajectory length for each English vowel in each group of children.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Bi-low M (SE)</th>
<th>Bi-high M (SE)</th>
<th>Mono M (SE)</th>
<th>p</th>
<th>$\eta^2_p$</th>
<th>Significant difference</th>
<th>Tukey HSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>499 (52)</td>
<td>300 (33)</td>
<td>254 (20)</td>
<td>&lt;0.001</td>
<td>0.56</td>
<td>Bi-high&lt;Bi-low, Mono&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>ɪ</td>
<td>268 (23)</td>
<td>253 (16)</td>
<td>181 (9)</td>
<td>0.002</td>
<td>0.456</td>
<td>Mono&lt;Bi-low, Mono&lt;Bi-high</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>345 (24)</td>
<td>459 (48)</td>
<td>457 (51)</td>
<td>0.169</td>
<td>0.156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ɛ</td>
<td>319 (33)</td>
<td>250 (17)</td>
<td>240 (14)</td>
<td>0.035</td>
<td>0.274</td>
<td>Mono&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>æ</td>
<td>352 (31)</td>
<td>274 (25)</td>
<td>247 (19)</td>
<td>0.022</td>
<td>0.306</td>
<td>Mono&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>325 (37)</td>
<td>595 (72)</td>
<td>517 (108)</td>
<td>0.106</td>
<td>0.193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ʊ</td>
<td>246 (32)</td>
<td>203 (18)</td>
<td>196 (15)</td>
<td>0.237</td>
<td>0.128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o</td>
<td>244 (18)</td>
<td>458 (36)</td>
<td>464 (32)</td>
<td>&lt;0.001</td>
<td>0.6</td>
<td>Bi-low&lt;Bi-high, Bi-low&lt;Mono</td>
<td></td>
</tr>
<tr>
<td>ɑ</td>
<td>318 (17)</td>
<td>295 (21)</td>
<td>240 (10)</td>
<td>0.007</td>
<td>0.379</td>
<td>Mono&lt;Bi-low</td>
<td></td>
</tr>
<tr>
<td>ʌ</td>
<td>282 (16)</td>
<td>251 (25)</td>
<td>264 (16)</td>
<td>0.563</td>
<td>0.053</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.11. Summary of statistical results showing the effect of L2 experience on trajectory length of each English vowel (means and standard errors are in Hz).
5.3.3 Shared vowels

Included among the five Mandarin vowels and ten English vowels are several pairs that have similar vowel quality and are shared by both vowel systems. In order to closely examine to what extent the L2 experience affect bilingual children’s production of these shared vowels, Mandarin /a, i, u/ and English /ɑ, i, u, æ/ were selected for further analysis. There was no single vowel matched with Mandarin /a/, therefore, two phonetically similar vowels /ɑ, æ/ were both chosen. Note that because each monolingual speaker just produced the words in their native language, the shared vowels of Mono group in this section are separately selected from monolingual Mandarin and monolingual English children.

5.3.3.1. Midpoint vowel dispersion of shared vowels

The mean of midpoint formant frequency values of each selected shared vowel for each group was plotted and shown in Figure 5.14. As shown in the first panel for Mono children, Mandarin /i/ and English /i/ were located close to each other while Mandarin /u/ and English /u/ showed a greater separation in the acoustic vowel space. For the vowel pairs of Mandarin /a/ and English /ɑ/, /æ/, Mandarin /a/ was located in a center and lower position relative to its two English counterparts. In addition, Mandarin /a/ is closer to English /æ/ than to English /ɑ/. In order to examine the extent to which the Mandarin and English similar vowels are similar or different, One-way ANOVA test was used on each
formant (F1 and F2) at midpoint for each vowel pair in each group of children. The results revealed no significant differences between Mandarin and English /i/ for either F1 or F2. However, for the Mandarin /u/ and English /u/, English /u/ showed a significant higher F2 (F(1, 22) = 138.445, \( p < 0.001, \eta_p^2 = 0.863 \)) than Mandarin /u/. This result demonstrated that English /u/ is more fronted than Mandarin /u/. For Mandarin /a/ and English /\textalpha/ and /æ/, there was a significant difference among these three vowels for both F1 (F(2, 30) = 11.208, \( p < 0.001, \eta_p^2 = 0.428 \)) and F2 (F(2, 30) = 60652, \( p < 0.001, \eta_p^2 = 0.802 \)). Tukey HSD results confirmed significant differences for F2 for each vowel pair (Mandarin /a/ - English /æ/, Mandarin /a/ - English /\textalpha/ and English /æ/- English /a/) and higher F1 in Mandarin /a/ than English /\textalpha/.

As shown in the middle panel for Bi-low children, similar to Mono children, Mandarin /i/ was also located close to English /i/. Again, the statistical results revealed no significant difference between these two /i/s. For the vowel pair of Mandarin /u/ vs. English /u/, similar to Mono children, English /u/ showed a significantly higher F2 (F(1, 12) = 39.563, \( p < 0.001, \eta_p^2 = 0.767 \)) than Mandarin /u/. However, as shown in the figure, English /u/ was located at a relative back position and was closer to Mandarin /u/ compared to that in Mono children. The Euclidean distance between the two /u/s in Bi-low children was only 616 Hz while that in Mono children was 1115 Hz which also
evidenced the approximation of Mandarin /u/ to English /u/ in Bi-low children. For Mandarin /a/ and English /a/ and /æ/, Bi-low children showed similar structure as those in Mono children. ANOVA tests revealed a significant difference for both F1 (F(2, 18) = 6.074, p < 0.001, \( \eta^2_p = 0.403 \)) and F2 (F(2, 18) = 19.664, p < 0.001, \( \eta^2_p = 0.686 \)). In addition, Tukey HSD showed a significantly higher F1 in Mandarin /a/ than English /a/.

However, for the pairwise comparison of F2, Bi-low children showed no significant difference between Mandarin /a/ and English /æ/, different from what was found in Mono children.

As shown in the bottom panel for Bi-high children, Mandarin /i/ and English /i/ were completely overlapped and no significant formant differences were found between these two /i/s. Similar to Mandarin /u/ and English /u/ in Mono children, Bi-high children also showed a large separation between these two vowels in the acoustic space. ANOVA results showed a higher F2 for English /u/ (F(1, 14) = 112.787, p < 0.001, \( \eta^2_p = 0.890 \)). As for the vowel of Mandarin /a/ and English /æ/ and /a/, the statistical results showed a significant difference for both F1 (F(2, 21) = 3.579, p = 0.046, \( \eta^2_p = 0.254 \)) and F2(F (2, 21) = 35.810, p < 0.001, \( \eta^2_p = 0.773 \)). However, unlike what was found in Mono children, Tukey HSD showed a significant difference between English /æ/ and /a/ but no significant difference between Mandarin /a/ and English /a/ for F1. As shown in the figure, the structure of these three vowels in Bi-high children was quite different from
that in Mono children. Specifically, Mandarin /a/ was considerably raised in Bi-high children, which was located at a position with similar vowel height as English /ə/ and /æ/.

In general, the statistical analyses on the midpoint formant frequency values showed no difference between Mandarin /i/ and English /i/ for any group of children. It showed a significantly higher F2 in English /u/ than Mandarin /u/ across all three groups of children. However, for Mandarin /a/ and English /ə/ and /æ/, the statistical analyses revealed inconsistency among these groups of children. Both Bi-low and Bi-high children showed some differences from Mono children on the structure of these three vowels.

Figure 5.14. Mean of midpoint formant frequency values for Mandarin and English shared vowels in each group of children.
Figure 5.14 continued
5.3.3.2 Vowel spectral change of shared vowels

As discussed earlier, midpoint formant frequency values only characterize the acoustic feature of a small static portion over the vowel duration. It fails to provide further information about the nature of the spectral change. To better examine to what extent the two groups of bilingual children were different from monolingual children in their vowel spectral change of these shared vowels, the formant trajectories of these vowels were plotted for each group and shown in Figure 5.15.

In the panel of Mono children, it is clearly shown that all three Mandarin vowels displayed opposite formant movement direction than their English counterpart vowels. For example, the trajectory of Mandarin /i/ moved to a relatively high and front position while the trajectory of English /i/ moved to a relatively low and back position from the 20 to 80% point. In addition, the magnitude of formant movement of Mandarin /i/ was larger than that of English /i/. These differences in the vowel spectral change of pair did not show in their midpoint formant frequency values. In this case, although these two /i/ showed similar relative position in the acoustic space, they were still different in the pattern of formant movement.

For the Bi-low children (shown in the middle panel), while they maintained the formant movement pattern of their Mandarin (native) vowels, their movement direction of the English vowels were quite different from those of Mono children. In particular, the formant trajectories of their English vowels moved in the same direction as their Mandarin counterparts. These consistent formant movement directions between
Mandarin and English shared vowels in Bi-low children suggest that these children were assimilating the vowel spectral change of their native language to the new vowels in the second language even though some vowels are very similar in the relative position such as Mandarin /i/ and English /i/.

For the Bi-high children (shown in the lower panel), the direction of formant trajectories in both Mandarin and English vowels conformed to those of Mono children. However, their Mandarin vowels still showed differences from Mono children in either the magnitude of the trajectories or the relative positions of the vowels. Specifically, their Mandarin /i/ showed much less formant movement than that in Mono children. These findings indicate that Bi-high children’s Mandarin vowels were affected by their English and thus showed different formant movement pattern from Mono children.
Figure 5.15. Vowel dynamic spectral change of Mandarin and English shared vowels for each group of children. Larger size symbols represent 80% point.
To examine the difference of the magnitude of formant movement between Mandarin and English shared vowels, One-way ANOVA was conducted on the trajectory length for each vowel pair in each group. For Mono children, the results showed a significant difference of TL for Mandarin /a/ and English /æ/ and /a/ (F(2,32) = 14.672, p < 0.001, $\eta_p^2 = 0.494$). In particular, the TL of Mandarin /a/ was significantly longer than that of English /æ/ and /a/. The TL of Mandarin /i/ was also significantly longer than that of English /i/ (F(1,22) = 46.557, p < 0.001, $\eta_p^2 = 0.679$). However, there was no significant difference of TL between Mandarin /u/ and English /u/. These results demonstrated that even though the Mandarin /i/ and English /i/ are similar in their relative position in the acoustic space, they are still different in the features of spectral change.

For the Bi-low children, no significant differences were found between Mandarin /a/ and English /æ/ and /a/, nor between Mandarin /i/ and English /i/. However, they produced significantly longer TL for Mandarin /u/ than for English /u/ (F(1,12) = 11.636, $p = 0.005, \eta_p^2 = 0.492$). While for the Bi-high children, no significant differences were found for all shared vowel pairs. These results indicates that both groups of bilingual children differ from monolingual children in their spectral change pattern, which suggest the influence of language experience on their vowel properties in their L1 and L2.
5.4 Discussion and Summary

This study was designed to examine the influence of L1-L2 interaction on the structure of vowel systems in young sequential bilingual children. Through comparing the static and dynamic vowel acoustic features of two groups of 5-6 years old bilingual Mandarin-English children with those of age-matched corresponding monolingual children, we found that young bilingual children with low proficiency in English preserved the basic acoustic features of all Mandarin vowels and showed limited difference from monolingual Mandarin children in the organization of acoustic vowel space and spectral change patterns. However, their production of English vowels demonstrated characteristics of beginning second language learners. In contrast, the young bilingual children with high proficiency in English produced English vowels in a manner generally similar to monolingual English children. However, they still show some different acoustic features of their vowel production relative to the monolingual age-matched. In addition, their production of Mandarin vowels seems to have been affected by their experience with English and showed systematic differences from monolingual Mandarin children.

The dynamic interaction between L1 and L2 is always a core issue in bilingualism research. In the present study, we observed the interaction effect in both directions in young bilingual children. First, considering the influence of L1 on L2, consistent with our prediction, the Bi-low children showed a strong effect of L1 on their production of L2 vowels. Similar to adulthood L2 learners, these young sequential bilingual children also adopted the phonetic system of their native language as the base to build the new sound systems for the L2. Therefore, during the initial stage of their immersion to the L2, they
tended to transmit the L1 phonetic features to the L2 production. In particular, for the
acoustic measures of vowel duration and trajectory length, Bi-low children generally
produced English vowels with longer duration and formant trajectories compared to Bi-
high and Mono English children. If we compare these acoustic measures in monolingual
Mandarin and English speakers, it can be observed that Mandarin vowels are always
produced with longer duration and greater formant movement than English vowels. This
may be associated with the different syllable structures in Mandarin and English. Thus,
when Bi-low children produce English vowels, they transfer certain temporal and spectral
features of their native language to the English vowels.

The influence of L1 on L2 can be more clearly observed in the scatter plots and
spectral change patterns in Bi-low children. The overlapped English vowel clusters in Bi-
low children indicate that this group of children is likely to assimilate L2 vowels into
close L1 vowel categories. The variation in each English vowel category suggests that Bi-
low children had not established separate acoustic categories for English vowels. In terms
of the vowel spectral change, Bi-low children produced English shared vowels with the
same formant movement direction as their Mandarin counterparts but completely
different from how monolingual English children produced. These findings evidenced
that Bi-low children transferred the spectral change pattern in their native language to the
second language.

Unlike in the Bi-low children, much less influence of L1 on L2 can be observed in
Bi-high children. Bi-high children produced most English vowels in a native-like manner

14 Mandarin syllables have no more than four phonemes. In addition, Mandarin syllables have no consonant
at the syllable coda position except nasal ending /n/ and /ŋ/.
in both temporal and spectral features. However, the longer trajectory length of English /i/ in Bi-high children relative to Mono children demonstrated that they still showed differences from monolingual English children in certain aspect of the acoustic characteristics. Although Bi-high children produced English /i/ with similar trajectory lengths as Bi-low children, which may be associated with longer trajectory length of the acoustically similar Mandarin vowel /i/, we still lack solid evidence to show the influence of L1 on L2 in Bi-high children.

In terms of the influence of L2 on L1, consistent with our prediction, Bi-high children produced their native vowels differently from Mono mandarin children in both static and dynamic acoustic characteristics. In particular, Bi-high children produced Mandarin /a/, /i/, /y/ and /ɤ/ in a different manner than Mono Mandarin children. Among these four Mandarin vowels, /a/ and /i/ have phonetically similar English counterparts. We observed that Bi-high children raised Mandarin /a/ close to English /ɑ/ and /æ/. In addition, Bi-high children transferred the dynamic feature of English /i/ to this Mandarin counterpart and produced Mandarin /i/ with much shorter trajectory length compared to that in Mono Mandarin children. The positional variation of Mandarin /a/ to its English counterparts and transfer of vowel dynamics feature of English /i/ to Mandarin /i/ evidenced the assimilatory process associated with the effect of L2 to L1.

For the other two Mandarin vowels /y/ and /ɤ/ which have no phonetically similar English counterparts, Bi-high children also showed different acoustic properties from
Mono Mandarin children. In particular, Bi-high children showed forward positional change of Mandarin /y/ relative to Mono Mandarin children. In addition, the formant trajectory of their Mandarin /y/ was characterized by shorter formant movement. We assume that the fronted and static /y/ in Bi-high children may be affected by their English high front vowels. English has two high front vowels /i/ and /ɪ/ but no vowel in a relatively high center position. In addition, both /i/ and /ɪ/ are characterized by a small amount of formant movement. Therefore, Bi-high children may transfer the acoustic features of English high front vowels to this unique Mandarin vowel /y/.

Regarding the vowel /ɤ/, Bi-high children produced this vowel with less amount of spectral change and tended to move this vowel to a central position. This centralization is also observed in Bi-low children. It seems that both group of bilingual children demonstrated the tendency of centralization of /ɤ/. However, the underlying driving force of this movement remains unclear.

In contrast to Bi-high children, the influence of L2 on L1 can hardly be traced in Bi-low children since there was no statistical difference between them and monolingual Mandarin children in any acoustic measurement. Overall the effect of L2 on L1 is clearly evident in Bi-high children. In particular, both similar and dissimilar L1 vowels are affected. Unlike those studies which found the deviation of close L1 sound from L2 sounds to maintain the vowel contrast in L1-L2 vowel systems, the present study
provides evidences in support of the approximation of L1 sounds to close L2 sounds, consistent with the finding in Chang (2012).

The primary difference between Bi-low and Bi-high children is their English experience. The difference in English experience is not just related to the length of English learning. It is also related to the manner of language learning in young children. As proposed in the environmental theory (Jia & Aaronson 2003), language learners who arrive at the L2 region at a young age are generally immersed in a richer L2 environment that includes not only a larger quantity but also a higher quality of L2 exposure. In addition, early learners of L2 are also highly motivated to use L2 to interact with peers and the dominant society culture. Therefore, in the present study, only two and half years’ difference in English learning between Bi-low and Bi-high children caused significant differences in phonetic features of these two groups.

Overall, our findings suggest that when sequential bilingual children acquire a second language, they used their L1 phonetic system as the base which exerts great influence on the phonetic structure of L2 at the beginning stage. Meanwhile, due to the lack of established L2 phonetic system, their L1 phonetic system remains relatively stable and is hardly influenced by their L2. When the bilingual children exposed to English for a relatively long period of time, once the L2 phonetic systems is well established, plus the flexibility of children’s phonetic system, the established L2 leads to modification of acoustic properties in their L1. In such a case, the bilingual children are less likely to retain the monolingual ability in their native language while they developed a native-like ability in their L2. Therefore, during the process of L2 learning, due to the L1-L2 interaction effect and plasticity of children’s phonetic systems, early bilingual children
can obtain the monolingual-like ability in L2, but meanwhile, they will also modify their L1 sounds accordingly.
References


181


Addendum

Word list used to collect Mandarin production

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Pinyin</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>da4 xiang</td>
<td>elephant</td>
</tr>
<tr>
<td></td>
<td>da4 suan</td>
<td>garlic</td>
</tr>
<tr>
<td>i</td>
<td>pi2 qiu</td>
<td>ball</td>
</tr>
<tr>
<td></td>
<td>bi2 zi</td>
<td>nose</td>
</tr>
<tr>
<td>u</td>
<td>tu4 zi</td>
<td>rabbit</td>
</tr>
<tr>
<td></td>
<td>pu2 tao</td>
<td>grape</td>
</tr>
<tr>
<td>y</td>
<td>ju2 zi</td>
<td>orange</td>
</tr>
<tr>
<td></td>
<td>yu2 tou</td>
<td>fish head</td>
</tr>
<tr>
<td>γ</td>
<td>ge1 ge</td>
<td>older brother</td>
</tr>
<tr>
<td></td>
<td>ge1 zi</td>
<td>pigeon</td>
</tr>
</tbody>
</table>
Word list used to collect English production

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>feet</td>
</tr>
<tr>
<td>i</td>
<td>kiss</td>
</tr>
<tr>
<td>e</td>
<td>cake</td>
</tr>
<tr>
<td>ε</td>
<td>desk</td>
</tr>
<tr>
<td>æ</td>
<td>cat</td>
</tr>
<tr>
<td>u</td>
<td>boot</td>
</tr>
<tr>
<td>ʊ</td>
<td>book</td>
</tr>
<tr>
<td>o</td>
<td>coat</td>
</tr>
<tr>
<td>θ</td>
<td>box</td>
</tr>
<tr>
<td>Λ</td>
<td>cup</td>
</tr>
<tr>
<td>ae</td>
<td>geese</td>
</tr>
<tr>
<td>kitchen</td>
<td></td>
</tr>
<tr>
<td>tape</td>
<td></td>
</tr>
<tr>
<td>pet</td>
<td></td>
</tr>
<tr>
<td>bat</td>
<td></td>
</tr>
<tr>
<td>goose</td>
<td></td>
</tr>
<tr>
<td>cookie</td>
<td></td>
</tr>
<tr>
<td>boat</td>
<td></td>
</tr>
<tr>
<td>stop</td>
<td></td>
</tr>
<tr>
<td>duck</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6: General Discussion and Summary

This dissertation has two important goals. First, we aim to investigate the acoustic development of vowels in children aged 3-6 years from monolingual Mandarin-and monolingual English-speaking environments. Second, we aim to document the developmental path of language separation and explore the effect of L1-L2 interactions on the acoustic features of vowels in young Mandarin-English bilingual children. This dissertation expands previous acoustic studies of vowel production in young children by systematically investigating both static and dynamic acoustic features of vowel production in both monolingual and bilingual children. The major findings can be broadly summarized in the following four aspects:

1. As shown in Chapters 2 and 3, vowels produced by monolingual Mandarin and monolingual English children at 3 to 6 years of age still show different temporal and spectral characteristics from those of adults. First, children generally produce longer vowel duration than adults. Second, the decreasing acoustic category sizes from children to adults observed in both Mandarin and English children indicate the continuing refinement of acoustic characteristics in children in this age range. Third, both Mandarin and English children show comparable patterns of formant movements to those of adults for most of vowels in their native languages, however, the magnitudes of formant movement for some vowels and the relative positions of
certain vowels are still different from the adult-norms, which indicates that children at this age have not developed adult-like features of spectral change.

2. The development path of L2 vowels in the emergent bilingual child in Chapter 3 and the acoustic properties of vowel production in Bi-low children relative to age-matched monolingual children in Chapter 4 support the one system stage in the speech development of early sequential bilingual children. Similar to adult L2 learners, early sequential bilingual children adopt their native language as the base for the acquisition of the second language system. The Mandarin-like triangular vowel space of English vowel system in the emergent bilingual child at the beginning of the recording period (shown in Chapter 3) and the substantial overlapped vowel categories in the bilingual children with low levels of English exposure (shown in Chapter 4) suggest that early bilingual children do transfer L1 to L2 at the beginning of L2 acquisition. However, unlike adult L2 learners, early sequential bilingual children complete the reorganization of the L2 vowel system within a short time span. As shown in this dissertation, the emergent bilingual child in Chapter 3 developed a near native-like vowel space within 20 months while those early bilingual children in Chapter 4 demonstrated a native-like ability in producing both static and dynamic vowel features after three years of immersion in English.

3. When comparing the development path of L2 vowels in the emergent bilingual child (Chapter 3) with the acoustic development of monolingual English children (Chapter 2), we found that the bilingual child acquired his English vowels in a manner different from the monolingual English children. It appeared that the bilingual child needed to restructure the overlapped English vowels resulted from the influence of L1
and meanwhile refine the acoustic features of the English vowels. This finding suggests that bilingual children may go through a different developmental path of L2 vowels than do young native learners of the language. On the other hand, bilingual children may also share some commonalities with monolingual children in the speech development. For example, the reduction in the overlap of the back vowel categories from younger English children to older English children (Chapter 2) and the continuing separation of back vowels in the emergent bilingual child by the end of recording period (Chapter 3) suggest that back vowels may develop later than front vowels in both monolingual and bilingual children. This may be associated with the more complex neuromuscular controls involved in producing back vowels.

Regardless of the language backgrounds, normal developing children at same age may show similar patterns of neuromuscular development and thus, monolingual and bilingual children demonstrate certain commonalities in their speech development.

4. Chapter 4 provides evidence to support the bi-directional interaction effect between L1 and L2 in young bilingual children. However, the strength and size of the bi-directional effect are determined by different stage of L2 learning in the bilingual children. When bilingual children were at the initial stage of L2 learning with low level of L2 proficiency (as shown in Bi-low children), they assimilated the L2 to L1 vowel categories and transferred the acoustic features to L2 vowels, similar to what adult L2 learners do. Meanwhile, they maintained the acoustic features of their L1 vowels and showed no clear effects of L2 on L1. Later, because of greater exposure to the L2 (as shown in Bi-high children), the bilingual children had established a native-like L2 vowel system. The L2 vowel features tended to affect the L1 vowels.
In particular, the L2 to L1 influence was demonstrated on both similar and dissimilar vowels. More importantly, the interaction effect was demonstrated not only on the static acoustic features but also on the dynamic spectral changes.

In summary, the acoustic analysis of monolingual Mandarin and English children’s vowel productions shows that the acoustic development of vowels in childhood speech is a long-term process, regardless of the inventory size of the vowel system. Children generally start from more scattered and less-distinguishable vowel categories and gradually tune their acoustic features to the adult targets as a function of chronological age and experience with the language. The acoustic development of bilingual children’s vowel production provided evidence on the plasticity of children’s speech sound system. In particular, young children can generally develop an adult-like L2 system in a relatively short period of time. Meanwhile, the speech sounds of their native language are more vulnerable to the influence of a second language and the effect of their native language on the second language acquisition is also less firm.
Bibliography


196


