Research Article

Jing Yang*

Vowel development in young Mandarin-English bilingual children

https://doi.org/10.1515/phon-2021-2006
Published online June 22, 2021

Abstract: This study examined the development of vowel categories in young Mandarin-English bilingual children. The participants included 35 children aged between 3 and 4 years old (15 Mandarin-English bilinguals, six English monolinguals, and 14 Mandarin monolinguals). The bilingual children were divided into two groups: one group had a shorter duration (<1 year) of intensive immersion in English (Bi-low group) and one group had a longer duration (>1 year) of intensive immersion in English (Bi-high group). The participants were recorded producing one list of Mandarin words containing the vowels /a, i, y, ɤ/ and/or one list of English words containing the vowels /i, ɪ, e, ɛ, æ, u, ʊ, o, ʌ/. Formant frequency values were extracted at five equidistant time locations (the 20–35–50–65–80% point) over the course of vowel duration. Cross-language and within-language comparisons were conducted on the midpoint formant values and formant trajectories. The results showed that children in the Bi-low group produced their English vowels into clusters and showed positional deviations from the monolingual targets. However, they maintained the phonetic features of their native vowel sounds well and mainly used an assimilatory process to organize the vowel systems. Children in the Bi-high group separated their English vowels well. They used both assimilatory and dissimilatory processes to construct and refine the two vowel systems. These bilingual children approximated monolingual English children to a better extent than the children in the Bi-low group. However, when compared to the monolingual peers, they demonstrated observable deviations in both L1 and L2.

Keywords: acoustic characteristics; Mandarin-English bilingual children; vowel development

*Corresponding author: Jing Yang, Department of Communication Sciences and Disorders, University of Wisconsin-Milwaukee, P.O Box 413, Enderis Hall 873, Milwaukee, WI 53201, USA, E-mail: jyang888@uwm.edu
1 Introduction

Phonological acquisition in bilingual children has raised an increasing amount of research interest in the past few decades. Many investigations have been carried out to document the acquisition order and error patterns of phonological systems in young bilingual children from diverse language backgrounds (e.g. Anderson 2004; Fabiano-Smith and Goldstein 2010; Goldstein and Washington 2001; Grech and Dodd 2008; Holm and Dodd 1999; Keshavarz and Ingram 2002). Only a mere handful of studies have examined the formation of phonetic categories and the development of acoustic characteristics in young bilingual children (Khattab 2000; Lee and Iverson 2012, 2017; Simon 2010; Turner et al. 2014; Yang et al. 2015). The phonetic aspect of bilingualism is of particular interest because it “constitutes a privileged window of linguistic inter-influence” (Sankoff 2001). The language change caused by language contact or second language acquisition is normally initially reflected in the speaker’s pronunciation and lexicon. In the present study, we conducted an acoustic-phonetic analysis to characterize vowel development at the early stage of language acquisition in young bilingual Mandarin-English children.

1.1 Theoretical framework

Researchers have proposed different theories and models to explain the development of phonetic categories in perception and production in bilingual children and second language learners (e.g. Perceptual Assimilation Model-L2, PAM-L2, Best and Tyler 2007; Speech Learning Model, SLM, Flege 1995; Second-Language Linguistic Perception model, L2LP, Escudero 2005; Processing Rich Information from Multidimensional Interactive Representations, PRIMIR, Curtin et al. 2011). Among these theories, PRIMIR addresses the development of speech perception, which has been recently extended to children growing up in bilingual or multilingual environments. SLM focuses on the formation and development of L2 sounds in speech production. These are directly related to our current data. PRIMIR assumes that monolingual and bilingual infants are equipped with the same representational spaces and statistical learning mechanisms. SLM posits that language learning capacity remains intact over the course of the whole life span. These assumptions ensure that bilingual children and even adult L2 learners possess the ability to develop monolingual-like speech and language abilities in both languages.
While the ultimate achievement of monolingual-like speech ability is assumed, the development process varies in children and is affected by multiple factors. PRIMIR emphasizes that compared to monolingual infants, bilingual infants receive different speech inputs and experience different task demands. In addition to the external factors, infants show internal perceptual bias and different approaches to process speech intake. These experiential influences of speech intake are reflected in both perception and production (Polka et al. 2007). These notions imply that in children, especially bilingual children, the developmental paths of their two languages may be different from the monolingual peers. Meanwhile, PRIMIR proposes that to keep pace with monolingual development, bilingual infants may utilize a comparison-contrast mechanism to facilitate language separation and development. Specifically, bilingual infants may compare the phonetic features between the two languages to group similar information and contrast different information to separate the two systems.

Whereas the comparison-contrast process implies the possibility of mutual influence between the two language systems, SLM explicitly proposes that the phonetic elements of L1 and L2 exist in a “common phonological space” in which the two phonetic systems can interact with each other. Moreover, this model provides learnability predictions and proposes assimilatory and dissimilatory mechanisms to explain phonetic changes in L2 development. The learnability of L2 sound categories is based on the magnitude of phonetic similarity between the L2 and L1 sounds. The easiest L2 sound is the one that completely overlaps with an L1 sound. The most difficult L2 sound is the one that shows phonetic-acoustic similarities but is not identical to the L1 sounds. For an L2 sound that has no overlap with an L1 sound, it will be treated as a new sound that requires substantial experience and practice in the L2 to be established.

During the process of speech development, there are two mechanisms to account for the phonetic change that may happen in both L1 and L2 sounds. Assimilatory movement refers to the approximation of two phonetic categories. Adult L2 learners at the beginning stage usually use similar L1 sounds to substitute or assimilate L2 sounds. In some experienced L2 learners, especially early bilingual speakers who have acquired the L2 phonetic system, assimilation of L1 sounds to L2 sounds may also occur (e.g. Chang 2012; Flege 1987; Kartushina et al. 2016). In contrast, dissimilatory process refers to the phonetic movement of shifting two similar sounds away from each other. This normally occurs when early bilingual speakers or experienced L2 learners move the L1 vowels from newly-established L2 vowels to maximize the L1-L2 distinction. It may also occur in beginning L2 learners who produce an L2 sound away from its target position or similar L1 category to exaggerate the phonetic contrast between the similar L1 and L2 sounds (e.g. Flege et al. 2003; Guion 2003). Although these two processes were initially
proposed to account for phonetic changes in adult L2 learners, they also occur in children’s speech development (Lee and Iverson 2012; Yang and Fox 2017).

1.2 Separation of phonetic systems in bilingual children

For children who learn a new language at a young age, the learning process typically proceeds in a quick manner. The onset of speech development, especially the L2 development in bilingual children, is tied to the long-lasting debate of “one system or two”. Some early work (Leopold 1949; Oksaar 1971; Redlinger and Park 1980; Schnitzer and Krasinski 1994; Volterra and Taeschner 1978) have suggested that bilingual children who are exposed to two languages since birth might go through a single merged system stage in their language development. By contrast, many studies (Genesee 1989; Johnson and Lancaster 1998; Johnson and Wilson 2002; Keshavarz and Ingram 2002; Meisel 1989; Paradis 2001; Schnitzer and Krasinski 1996) have reported that bilingual children can separate their two languages from the earliest stage of language acquisition, which is primarily manifested in the phonological domain. As for the bilingual children who learn the two languages in successive order, Watson (1991) proposed that they first utilize the L1 as the foundation and superimpose the L2 on the L1. Then, they begin a process of separating L2 from L1 by adding or modifying the components in the L2.

With increasing exposure to L2, bilingual children experience continuing development of the new language. During this process, the two systems interact with each other and show an interdependence model (Fabiano-Smith and Barlow 2010; Johnson and Lancaster 1998). Recently, there has been an increasing body of research examining L1–L2 interactions during the speech development in bilingual children (Baker and Trofimovich 2005; Baker et al. 2008; Darcy and Krüger 2012; Lee and Iverson 2012, 2017; Oh et al. 2011; Yang and Fox 2017; Yang et al. 2015). These studies have shown that bilingual children and early L2 learners are more likely to show a dynamic interaction between L1 and L2 than adult L2 learners (Baker and Trofimovich 2005; Guion 2003; Yang and Fox 2017). Baker and Trofimovich (2005) demonstrated how L1 (Korean) and L2 (English) vowel systems interacted in bilinguals who differed in the age of exposure to L2 (older children and young adults) and the amount of experience with L2 (1 and 7 years). 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. By contrast, for older children, particularly those who had used their L2 for seven years, phonetic restructuring as a function of L2 learning affected both languages. The L1 and L2 vowels produced by the bilingual children with extended L2 use were
more susceptible to bidirectional influences. These findings suggest that L1 and L2 interact differently depending on the age at which L2 is learned. Children may separate the L1 and L2 in a different manner than adults do. In a more recent study by Yang and Fox (2017), the authors compared the static and dynamic vowel spectral features in two groups of 5–6-year-old bilingual Mandarin – English children and two groups of age-matched monolingual children. The results indicated that the bilingual children with low proficiency in English showed a strong influence of L1 on L2, while the bilingual children with high proficiency in English showed an influence of L2 to L1. These L1–L2 interactions were manifested in both static and dynamic spectral changes.

In addition to the dynamic interactions between the two phonetic systems, different types of assimilatory and dissimilatory movements have been observed in bilingual children of different ages. Lee and Iverson (2009, 2012, 2017) examined how these adaptive mechanisms were used by 3-, 5- and 10-year-old bilingual Korean-English children to restructure the stop and vowel systems in both languages. Their results showed that 10-year-olds employed both assimilatory (merging Korean /ʌ/ with English /ɔ/) and dissimilatory mechanisms (producing English /æ/ with a higher F1 than monolingual English speakers). The five-year-old bilingual children, however, had not fully separated the stop systems in these two languages and mainly showed assimilatory changes of stop features. Different from the 5 and 10-year-olds, the three-year-old bilingual children produced stops and vowels similarly to the monolingual children and showed limited interaction between L1 and L2 phonetic systems. Oh et al. (2011) studied vowel productions in Japanese –English bilingual children who started to learn the L2 older than five in comparison to corresponding monolingual children. Their data showed an assimilatory shift of Japanese and English /i/ and two types of dissimilatory shifts. In particular, the Japanese /a/ and English /ʌ/ overlapped at the beginning of the study. After one year, the bilingual children demonstrated a fronted movement of Japanese /a/ (representing dissimilation of L1 from L2) and further back movement of English /ʌ/ (representing dissimilation of L2 from L1). Inconsistent with these studies, Yang and Fox (2017) found assimilatory movements of L2 vowels to L1 counterparts in five-year-old bilingual Mandarin-English children with low proficiency in English and assimilatory movements of L1 vowels to L2 counterparts in bilingual children with high proficiency in English.

1.3 Comparison between bilinguals and monolinguals

Deriving from the interaction account, a relevant question is whether bilingual children perform as monolingual peers or still show differences from the
monolinguals. This question has been addressed in many studies but no consensus has been reached. Some studies reported that young bilingual children and early bilingual adults who started to acquire an L2 before puberty showed little or no difference in the production and perception of L2 sounds from the native speakers (Flege et al. 1999; Mack 1989; MacLeod and Stoel-Gammon 2005; Lee and Iverson 2012, 2017; Oh et al. 2011). In Lee and Iverson (2012, 2017) studies on vowel development in 3-, 5-, and 10-year-old bilingual Korean-English children, the bilingual children in all age groups produced distinctive vowel categories across languages and produced formant frequencies similarly to monolingual children. Oh et al. (2011) found that the bilingual children showed significant changes in their formant values over the one-year study period. At the end of the study period, the bilingual children showed no statistical difference from age-matched native-English children. 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 (e.g. MacLeod et al. 2009).

Diverging from studies showing no differences between bilinguals and monolinguals, many studies revealed that while bilinguals could use various features to separate the two phonetic systems, they might still differ from monolinguals in certain aspects of their production and perception (Højen and Flege 2006; Pallier et al. 1997; Sebastian-Galles and Soto-Faraco 1999; Tsukada et al. 2005). Khattab (2000) compared VOT in three English-Arabic bilingual children aged between 5 and 10 years old and three monolingual children in each language. The bilingual children developed distinct VOT systems for the two languages, but the VOTs in each language were still different from those in the monolingual children. Kehoe (2002) recorded speech samples in three German-Spanish bilingual children and monolingual children of each language at three time-points over a two-year period, which started when the children were at the age of 1;7 (years;months). While the bilingual children showed no differences from the monolinguals in Spanish, they showed later acquisition in vowel length contrast in German in comparison to the monolingual controls. Darcy and Krüger (2012) compared the acoustic characteristics of selected vowels between bilingual Turkish – German children and monolingual German children. The bilingual children aged between 9 and 12 years old and started to learn German before three years of age. The authors found that these bilingual children employed both spectral and durational features to separate L1 and L2 vowels in their production, produced appropriate durational difference, and made clear tense-lax distinctions for the L2 vowels. However, they still demonstrated slight positional deviance in the vowel space for certain German vowels from the monolingual speakers.
1.4 Present study

The target languages in the present study include Mandarin and English that differ in a systematic way in the vowel systems. American English has a relatively large vowel inventory with at least 14 monophthongal vowel phonemes /i, ɪ, e, ɛ, æ, u, ʊ, o, ɔ, a, ə, ɚ, ɝ/ (Ladeforged 2001) while Mandarin has a small vowel inventory with only five monophthongal vowel phonemes /a, i, u, y, ɤ/ (Duanmu 2007). Second, English is characterized by tense-lax vowel contrasts, which do not occur in Mandarin. Third, the English vowel system presents a quadrilateral vowel space with the four vowels /i, æ, ɑ, u/ occupying the corner positions. By contrast, the Mandarin vowel system is structured in a triangular shape with the three vowels /a, i, u/ occupying the corner positions. To date, a number of phonetic studies have been undertaken to examine the vowel production in Mandarin-speaking adults who learned English as an L2 (Chen 2006; Chen et al. 2001; Jiang 2008; Liao 2006; Wang 1997). Little research has been done to document the acoustic development and formation of phonetic categories in young bilingual Mandarin-English children. Yang et al. (2015) documented the acoustic development of L2 in one Mandarin-speaking child who learned English at a young age of 3;7. The vowel productions of this child were recorded on a regular basis over a 20-month period, and the first recording was implemented two weeks after the child first enrolled in an English daycare. The data revealed that the child first adopted the L1 phonetic categories as the foundation to build the vowel system for the L2. Then, the child went through a drastic restructuring and refining phase to separate the L2 vowel categories from the L1 system, which led to monolingual-like vowel productions for the L2 at the end of the study period.

While this single-subject longitudinal study enabled us to track the process of separation and development of vowel systems in a particular participant, a group study with more participants would be necessary to validate the findings. The present study recruited a group of young bilingual Mandarin (L1)-English (L2) children aged between 3 and 4 years old, with varying lengths of immersion in English. The goal of the present study is three-fold:

1. To examine whether young bilingual children, at the early stage of L2 development, can successfully separate the two phonetic systems and vowel categories in L2;
2. To examine how young bilingual children use assimilatory and dissimilatory processes to organize the two phonetic systems;

---

1 The present study chose the symbol /ɤ/ instead of /a/ used in Duanmu (2007) to represent the high-mid, back, unrounded vowel as a monophthong.
3. To identify whether and to what extent young bilingual children differ from the corresponding monolingual peers in both L1 and L2.

2 Methods

2.1 Participants

The participants of the present study included 35 3- and 4-year-old children. Specifically, there were 15 bilingual Mandarin-English children, 14 monolingual Mandarin children, and six monolingual English children. All monolingual Mandarin children (eight males, six females; Mean age = 3;11) were recruited from a kindergarten located within the Beijing area in China. These children were raised in a Mandarin-speaking environment with both parents speaking Mandarin in their daily life. The monolingual English children (two males, four females; Mean age = 3;6) and bilingual children were recruited by personal invitation in the Midwest region of the U.S. The monolingual English children were born and raised in an English-speaking environment with both parents as monolingual English speakers. All of the bilingual children were raised in a Mandarin-speaking environment with Mandarin-speaking parents. These bilingual children learned Mandarin since birth as their first/primary language. Table 1 shows the detailed demographic information and language usage of each child. Eleven of the 15 bilingual children were born in the U.S., and the other four arrived in the U.S. at various ages. All bilingual children had a very limited amount of exposure to English before being enrolled in the English daycare or kindergarten when the intensive English exposure and formal English learning started. Note that two bilingual children started immersion in English at a very young age, and another two bilingual children enrolled in a part-time English daycare program. Based on the length of intensive immersion in English, the bilingual children were divided into two subgroups: Bi-low (Mean age = 3;7) and Bi-high (Mean age = 4;2). Bi-low children had a shorter duration of exposure to English (<1 year), and Bi-high children had a longer duration of exposure to English (>1 year). Correspondingly, the Bi-low children had a relatively low percentage of daily use in English while the Bi-high children had a relatively high percentage of English use in their daily life. All monolingual and bilingual children were reported as having no history of language, speech, or hearing problems.
Table 1: Demographic information and language usage of the bilingual Mandarin-English children. AOL, Age of learning; LOR, Length of residence; LOE, Length of exposure.

<table>
<thead>
<tr>
<th>Group</th>
<th>Subj</th>
<th>Gender</th>
<th>Age (yrs;mos)</th>
<th>AOL_E (yrs;mos)</th>
<th>LOR_US (yrs)</th>
<th>LOE_E (^a) (yrs)</th>
<th>Daily_use_E</th>
<th>AOL_M</th>
<th>LOR_China (yrs)</th>
<th>Daily_use_M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-low</td>
<td>1</td>
<td>M</td>
<td>3;1</td>
<td>1;8(^a)</td>
<td>2.67</td>
<td>0.62</td>
<td>50%</td>
<td>Since birth</td>
<td>0.42</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>M</td>
<td>3;2</td>
<td>2;6</td>
<td>3.17</td>
<td>0.67</td>
<td>30%</td>
<td>Since birth</td>
<td>0.00</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>M</td>
<td>3;5</td>
<td>3;4</td>
<td>1.17</td>
<td>0.09</td>
<td>&lt;30%</td>
<td>Since birth</td>
<td>2.25</td>
<td>&gt;70%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>M</td>
<td>3;8</td>
<td>2;9</td>
<td>1.25</td>
<td>0.92</td>
<td>50%</td>
<td>Since birth</td>
<td>2.42</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>M</td>
<td>3;10</td>
<td>3;6</td>
<td>0.33</td>
<td>0.33</td>
<td>50%</td>
<td>Since birth</td>
<td>3.50</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>M</td>
<td>4;0</td>
<td>3;6</td>
<td>4.00</td>
<td>0.5</td>
<td>30%</td>
<td>Since birth</td>
<td>0.00</td>
<td>70%</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>F</td>
<td>4;1</td>
<td>2;11(^b)</td>
<td>2.50</td>
<td>0.35</td>
<td>30%</td>
<td>Since birth</td>
<td>1.58</td>
<td>70%</td>
</tr>
<tr>
<td>Bi-high</td>
<td>8</td>
<td>F</td>
<td>3;0</td>
<td>1;8</td>
<td>3.00</td>
<td>1.33</td>
<td>70%</td>
<td>Since birth</td>
<td>0.00</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>M</td>
<td>4;1</td>
<td>2;9</td>
<td>1.75(^c)</td>
<td>1.33</td>
<td>50%</td>
<td>Since birth</td>
<td>0.33</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>F</td>
<td>4;1</td>
<td>0;4</td>
<td>3.75</td>
<td>3.75</td>
<td>90%</td>
<td>Since birth</td>
<td>0.33</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>F</td>
<td>4;1</td>
<td>0;6</td>
<td>4.08</td>
<td>3.58</td>
<td>70%</td>
<td>Since birth</td>
<td>0.00</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>M</td>
<td>4;5</td>
<td>3;1</td>
<td>4.42</td>
<td>1.34</td>
<td>50%</td>
<td>Since birth</td>
<td>0.00</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>M</td>
<td>4;6</td>
<td>3;3</td>
<td>4.50</td>
<td>1.25</td>
<td>50%</td>
<td>Since birth</td>
<td>0.00</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>F</td>
<td>4;7</td>
<td>2;0</td>
<td>4.58</td>
<td>2.58</td>
<td>&gt;70%</td>
<td>Since birth</td>
<td>0.00</td>
<td>&lt;30%</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>M</td>
<td>4;8</td>
<td>2;0</td>
<td>4.67</td>
<td>2.67</td>
<td>&gt;70%</td>
<td>Since birth</td>
<td>0.00</td>
<td>&lt;30%</td>
</tr>
</tbody>
</table>

\(^a\)This child spent two half-days per week in the daycare till 2.8; \(^b\)This child spent three half-days per week in the daycare; \(^c\)This child was born and had lived in Denmark before moving to the U.S.; \(^d\)LOE_E was calculated by subtracting AOL_E from Age. For the two participants who were enrolled in a part-time program, the LOE_E was converted based on how much time they spent in the English daycare with reference to a full-time program (number of half-days/10).
2.2 Speech materials

Table 2 shows the word lists used to elicit vowel productions from the monolingual and bilingual children. The Mandarin word list included 10 disyllabic words containing the five monophthongal vowel phonemes /a, i, u, y, ɤ/. The target vowels always occurred in the first CV syllable. Stops were selected as the initial consonant except for the words containing the vowel /y/. Due to the phonotactic constraints in Mandarin, the vowel /y/ was presented with no initial consonant or with an affricate /ʨ/. The tone environment was not strictly controlled, except that tone 3 was avoided to limit the tone-vowel interactions (Chang 2010). The English word list included 20 monosyllabic or disyllabic words containing the 10 English monophthongal vowel phonemes /i, ɪ, e, æ, u, o, ʌ, α/. Note that due to the merger of the low back vowels /ɔ/ and /ɑ/ in most dialects of American English, only vowel /ɑ/ was examined in the present study. For the disyllabic words, the target vowels always occurred in the first syllable. Similar to Mandarin words, stops were selected as the consonant preceding the target vowels, and voiceless stops or fricative /s/ were selected as the consonants following the target vowels. The vocabulary size of young age children and the picturability of the target words were taken into consideration when developing the word lists.

2.3 Recording

The monolingual Mandarin children were recorded in a quiet, small conference room in a kindergarten located in the Beijing area. The monolingual English children and bilingual children were recorded in a quiet room at the children’s

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Pinyin</th>
<th>IPA</th>
<th>Mandarin</th>
<th>IPA</th>
<th>English</th>
<th>Word</th>
<th>IPA</th>
<th>Word</th>
<th>IPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>dà xiàng</td>
<td>/ta ɕian/</td>
<td>elephant</td>
<td>i</td>
<td>feet</td>
<td>/fit/</td>
<td>geese</td>
<td>/ɡiːs/</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>pí qiú</td>
<td>/pʰtɕʰiou/</td>
<td>ball</td>
<td>e</td>
<td>cake</td>
<td>/kɛk/</td>
<td>tape</td>
<td>/tep/</td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>tǔ zi</td>
<td>/tʰu tsʰi/</td>
<td>rabbit</td>
<td>æ</td>
<td>cat</td>
<td>/kæt/</td>
<td>bat</td>
<td>/bæt/</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>yǔ tóu</td>
<td>/yʰtʰou/</td>
<td>fish head</td>
<td>o</td>
<td>coat</td>
<td>/kot/</td>
<td>boat</td>
<td>/bote/</td>
<td></td>
</tr>
<tr>
<td>yj</td>
<td>gé ge</td>
<td>/kʰv kʰ/</td>
<td>older brother</td>
<td>a</td>
<td>box</td>
<td>/boks/</td>
<td>stop</td>
<td>/stɒp/</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>gé zi</td>
<td>/kʰv tsʰi/</td>
<td>pigeon</td>
<td>ʌ</td>
<td>cup</td>
<td>/kʌp/</td>
<td>duck</td>
<td>/dʌk/</td>
<td></td>
</tr>
</tbody>
</table>
home. For both monolingual and bilingual children, a questionnaire containing questions regarding participants’ demographic information, history of language learning, and daily usage of the language(s) was completed by the parents before the recording started. During the recording, each participant was seated in front of a laptop computer wearing a Shure SM10A head-mounted microphone situated approximately 1-inch from the participant’s mouth. A visual-auditory word-repetition task, controlled by a custom MATLAB program, was used to record speech samples. For each language, the recording session consisted of two blocks. In each block, the target words were randomly ordered, and pictures showing the target words were presented on the computer screen. An audio prime was played to the participants immediately after the display of each picture, and the participants were required to repeat each target word once. The audio primes were produced by a young female native speaker of each language. All speech samples were directly recorded onto a hard drive disk with a 16 bit quantization rate and a 44.1 kHz sampling rate. Note that the monolingual children only produced the word list in their native language while the bilingual children produced the word lists in both languages with a 15–20 min break between the Mandarin and English recordings. Considering the young age of participants, a word-repetition task was used to ensure consistent presentations of the stimulus words and better control of target word elicitation (Edwards and Beckman 2008).

2.4 Acoustical analysis

Due to skipped recording, noise overlay, unmeasurable tokens, etc., a total of 64 tokens (approximately 4%) were missing for subsequent analysis. The acoustical analyses were implemented using the waveform editing software Adobe Audition 1.0 and the spectrographic analysis program TF32 (Milenkovic 2003). Prior to the acoustic measurement and formant tracking, all tokens were down-sampled to 11.025 kHz using a custom MATLAB program. Then, the landmark locations of vowel onset and offset were located manually on the basis of the waveform with a visual check of the spectrogram. Vowel onset was set at the start of periodicity with visible F1 and F2 traces. Vowel offset was set at the end of periodicity extending through F1 and F2. These measurements were used for the calculation of vowel duration and were used as the input for subsequent formant extraction. For each token, the Linear Predictive Coding (LPC) formant tracks were obtained using the time-frequency analysis program TF32. Hand correction was employed when any errors in formant estimation were found. The frequency values of the first two formants (F1 and F2) of the target vowel were extracted from the retrieved formant tracks at five equidistant temporal locations (the 20–35–50–65–80% point) using a
custom MATLAB program. The 20% and 80% points were selected as the endpoints for formant extraction to limit the coarticulation effect of the neighbor consonant on the formant values at the transition portions. All landmark location measurements and formant tracks were re-checked. Uncertain measurements and formant tracks were checked by a second analyst. Any discrepancies were solved by a discussion between the two analysts.

It is well known that vowel formant frequency values are influenced by the speakers’ vocal tract size that varies with speakers’ sex and chronological age. Therefore, vowel normalization is always necessary if formant values need to be compared between males and females and among speakers from different age groups. In the present study, the speakers were young children aged between 3 and 4 years old. Regarding the sex difference, previous studies suggest that the physiological change of vocal tract structure related to gender difference does not occur before seven years of age and the effects of vocal tract length on the formant values between sexes during the pre- and peri-puberty periods are small (Lee and Iverson 2009; Vorperian et al. 2005). Therefore, we consider that the formant values caused by sex differences were negligible. Regarding the formant change caused by different vocal tract sizes, as the age range of our participants was relatively narrow, the effect of vocal tract size on formant values should be limited (Vorperian and Kent 2007). However, considering that almost all participants in the Bi-high group were older than those in the Bi-low children, normalized formant values were used when conducting group comparisons between the monolingual and bilingual children within the same language. In this present study, Lobanov’s vowel-extrinsic normalization approach was used to convert the Hz values into Z scores (Lobanov 1971). It is noteworthy that while vowel-extrinsic normalization is effective in eliminating the formant value differences caused by vocal tract sizes, it introduces artifacts when languages or dialects containing different vowel systems are compared (Clopper 2009). Therefore, when conducting the cross-language comparison within each group of children, the raw formant values were used.

2.5 Vowel trajectory length

In addition to the conventional midpoint formant frequency values, the present study examined the dynamic spectral change of L1 and L2 vowels in bilingual children with reference to the monolingual controls. Previous studies have shown that vowels, even monophthongs, are characterized by a certain amount of inherent spectral change (Assmann et al. 2013; Harrington and Cassidy 1994). Vowel dynamics reflected in the features of direction, shape, length, rate of change, etc., contain information indicating the articulatory gestures and
movement of vowel productions. The present study extracted the frequency values of F1 and F2 at five equidistant time locations to capture the dynamic feature of vowel spectral change. On the basis of these formant values, we calculated the formant trajectory length (TL) that is defined as the sum of the Euclidian distances between each two consecutive time locations (Fox and Jacewicz 2009):

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

(1)

where the distance between each two consecutive points (VSL) is calculated based on the formula:

$$VSL_n = \sqrt{(F_{1n+1} - F_{1n})^2 + (F_{2n+1} - F_{2n})^2}$$

(2)

In the present study, we compared formant dynamics of similar vowels in L1 and L2 of bilingual children with reference to monolinguals to understand how different language inputs and L1–L2 interactions affect vowel characteristics in young bilingual children. For example, as stated earlier, Mandarin vowels do not have the tense-lax contrast. Previous literature has documented that English lax vowels are challenging for Mandarin-speaking adults (e.g. Chen 2006). We wondered whether the acoustic features of English lax vowels reflected on formant dynamics would be acquired by the bilingual children.

3 Results

3.1 Comparison between Mandarin and English in bilingual and monolingual children

Figure 1 shows the vowel dispersion based on midpoint formant frequency values in Mandarin and English in each group of children. As shown in the left panel for the monolingual children, Mandarin vowels were structured in a triangle vowel space while English vowels were organized in a quadrilateral vowel space. Mandarin /i/ and English /i/ demonstrated substantial positional proximity and highly overlapped in the vowel space. Mandarin /ɤ/ showed positional approximation to the two mid-high English vowels /ʊ/ and /o/. Interestingly, although English /u/ has a phoneme counterpart of Mandarin /u/, it was located in a fronted position approximating Mandarin /y/ but deviating from Mandarin /u/. As for the Bi-low children (middle panel), their Mandarin vowels were organized in a way similar to monolingual Mandarin children. However, their English vowels were clustered into several groups. Specifically, /ʌ/ shifted to a low position close to English /a/.
The vowel /æ/ moved up and merged with the vowel /ɛ/. The vowels /e/ and /ɪ/ moved close to each other. The vowel /u/ retracted back, and the vowels /ʊ/ and /o/ moved up and back, which resulted in positional approximation among the three vowels in this group. Unlike the Bi-low children, the Bi-high children showed widely separated English vowels. But they still showed slight differences from monolingual English children in that they produced English /u/ at a further back position and separated the English vowels /ʊ/ and /o/ to a greater extent than the monolingual English children did. Meanwhile, their Mandarin vowels showed some variations from the monolingual Mandarin children. The biggest change was the approximation of Mandarin /a/ towards English /ɑ/. In addition, Mandarin /y/ and /u/ produced by the Bi-high children demonstrated greater positional variation and Mandarin /ɤ/ in the Bi-high children highly overlapped with their English /ʊ/.

As shown in the comparison of Mandarin and English vowel space in monolingual children, the five Mandarin vowels presented positional approximation to certain English vowels. A MANOVA test was conducted to analyze the formant values of similar Mandarin-English vowel pairs. The purpose of this test was to examine how bilingual children treat similar vowels in these two languages. The tested Mandarin-English vowel pairs included M/i/-E/i/, M/u/-E/u/, M/a/-E/ɑ/, M/a/-E/ɤ/, M/y/-E/u/, M/y/-E/u/. Given that multiple comparisons were conducted for each group, a Bonferroni correction with an alpha level of 0.009 was used. The results (summarized in Table 3) revealed a significant multivariate effect for M/a/-E/ɤ/ and M/u/-E/u/ in the monolingual children, M/a/-E/ɤ/ and M/y/-E/u/ in the Bi-high group. As shown in Table 4, the univariate analysis on F1 and F2, respectively, suggest that Mandarin /a/ differed from English /ɤ/ in F1 for both monolingual and Bi-high children. Mandarin /u/ and English /u/ produced by monolingual children showed differences in F2. However, for the Bi-high group, Mandarin /y/ differed from English /u/ in F2. These results indicate that the Bi-high
children separated the L1 and L2 in a better way than the Bi-low children did, but
they still differed from monolingual English children in the organization of the two
vowel systems.

Figure 2 presents a cross-language comparison of the group mean formant
trajectories for Mandarin and English vowels in each group. As shown in the left
panel for the monolingual children, all Mandarin vowels were produced with a
certain amount of formant movement even though they are all monophthongs. The
two English vowels /e/ and /o/ showed the greatest magnitude of movement than
the other vowels, which indicates the diphthongization of these two monoph-
thongal phonemes. The direction of the formant trajectories reflects the gliding
movement from the first vowel target to the second target. By contrast, English lax
vowels /ɪ/ and /ʊ/ that are close to /e/ and /o/ in the vowel space showed a smaller
magnitude of movement. The shared vowels in the two languages (e.g. Mandarin
/i/ and English /i/) demonstrated positional approximation at the midpoint loca-
tions but differed in the direction of formant movement.

The Bi-low children produced Mandarin /i/ and English /i/ with a similar
magnitude but opposite direction of formant movement, as was shown in mono-
lingual comparison. The trajectories of the English vowels /u, ū, o/ in the high back
region were close to each other but still separable from the trajectories of Mandarin
/u/ and /ɤ/ in the general shape and direction of formant movement. The magnitude of formant movement of English lax vowel /ʊ/ was smaller than the
surrounding English and Mandarin vowels in a similar region. The trajectories of
the two English vowels /e/ and /æ/ in the front low region highly overlapped and
the two English vowels /ʌ/ and /ɑ/ in the low back region showed a similar
magnitude of formant movement with similar vowel height. The trajectory of

| Table 3: Summary of MANOVA results on formant values of Mandarin-English similar vowel pairs in the monolingual and bilingual children. |
|-----------------|-------|-----|-------|---|-------|-------|---|-------|-------|---|
|                | Wilk’s λ | F    | p     |     | Wilk’s λ | F    | p     |     | Wilk’s λ | F    | p     |
| M/ɑ/-E/ɑ/      | 0.761   | 2.672 | 0.098 |     | 0.651   | 2.95 | 0.094 |     | 0.586   | 4.594 | 0.031 |
| M/a/-E/ʌ/      | 0.571   | 6.380 | 0.009*|     | 0.825   | 1.167| 0.347 |     | 0.37    | 11.055| 0.002*|
| M/i/-E/i/      | 0.825   | 1.804 | 0.195 |     | 0.554   | 4.428| 0.039 |     | 0.687   | 2.96  | 0.087 |
| M/u/-E/u/      | 0.061   | 131.798 | <0.0001*| | 0.556   | 4.386| 0.04  |     | 0.611   | 4.139 | 0.041 |
| M/ɤ/-E/ʊ/      | 0.701   | 3.628 | 0.049 |     | 0.544   | 4.605| 0.035 |     | 0.482   | 6.992 | 0.009*|
| M/ɜ/-E/ʌ/      | 0.833   | 1.699 | 0.213 |     | 0.458   | 6.521| 0.014 |     | 0.993   | 0.045 | 0.956 |

*Represents significant results after Bonferroni adjustments for multiple comparisons.
Table 4: Summary of univariate results on F1 and F2 of Mandarin and English similar vowel pairs for the monolingual and bilingual children.

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bi-low</th>
<th>Bi-high</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1</td>
<td></td>
<td>F1</td>
</tr>
<tr>
<td>M/a-/E/a/</td>
<td>0.194</td>
<td>0.092</td>
<td>4.832</td>
</tr>
<tr>
<td>M/a-/E/ʌ/</td>
<td>8.861</td>
<td>0.008*</td>
<td>1.69</td>
</tr>
<tr>
<td>M/i-/E/i/</td>
<td>2.247</td>
<td>0.151</td>
<td>0.022</td>
</tr>
<tr>
<td>M/u-/E/u/</td>
<td>1.145</td>
<td>0.299</td>
<td>178.885</td>
</tr>
<tr>
<td>M/y-/E/u/</td>
<td>4.676</td>
<td>0.044</td>
<td>4.966</td>
</tr>
<tr>
<td>M/y-/E/ʊ/</td>
<td>0.16</td>
<td>0.693</td>
<td>3.486</td>
</tr>
</tbody>
</table>

*Represents significant results after Bonferroni adjustments for multiple comparisons. *This is not marked as significant because the MANOVA test revealed no significant difference (with Bonferroni adjustments) for this vowel pair in the Bi-low children.
Mandarin /a/ was close to English /ʌ/ in the relative position but differed from /ʌ/ in the direction of formant movement. The Bi-high children produced the trajectories of Mandarin /i/ and English /i/ similarly to the monolingual and the Bi-low children. The trajectories of English /u/ approximated Mandarin /u/ in the Bi-high children, as was the case for the Bi-low children but different from the monolingual children. Mandarin /ɤ/ and English /ʊ/ were located in a similar position in the Bi-high group. The trajectories of these two vowels were both different from the corresponding vowels produced by monolingual children in the shape, direction, and magnitude features.

The formant TL was compared between Mandarin and English similar vowels in monolingual children and two groups of bilingual children. The same six Mandarin-English vowel pairs were examined. An independent sample t-test was used for the vowels produced by the monolingual children and a paired-wise sample t-test was used for the vowels produced by the bilingual children (results summarized in Table 5). A Bonferroni correction with an alpha level of 0.009 was used to accommodate multiple comparisons. For monolingual children, the TL of Mandarin /a/ was significantly different from the TL of English /ʌ/ and the TL of Mandarin /ɤ/ was significantly different from the TL of English /ʊ/. While similar magnitude differences seem to be present between Mandarin /ɤ/ and English /ʊ/ as well as Mandarin /a/ and English /ʌ/ in the two bilingual groups, no significant difference was yielded after the Bonferroni adjustment. But there was a significant difference between Mandarin /a/ and English /ʌ/ in the Bi-high children.

Figure 2: Formant trajectories showing the vowel spectral change of Mandarin and English vowels in the two groups of bilingual children and corresponding monolingual children. The trajectories were plotted based on the formant frequencies at five time locations (the 20–35–50–65–80% point) over the course of vowel duration. The larger size symbols represent the 80% point.
3.2 Comparison between bilingual and monolingual children in Mandarin and English

To better examine the phonetic shift caused by the L1–L2 interaction and identify to what extent the vowel sounds in the L1 and L2 differed from those by the monolingual children, normalized formant frequency values were compared across three groups of children for Mandarin and English, respectively (shown in Figure 3). Both Bi-low and Bi-high children produced the Mandarin vowels, except for the vowel /y/, similarly to monolingual Mandarin children. The two bilingual groups produced Mandarin /y/ with a lower vowel height and demonstrated greater variations in F2. As for the English vowels, both groups of bilingual children demonstrated positional deviations from the monolingual English children but the Bi-high children approximated the monolingual peers to a greater extent than the Bi-low children did. Among the 10 tested English vowels, the Bi-low children approximated monolingual targets for the three vowels /i, ɪ, e/ located in the high front region. For the three vowels /u, ʊ, o/ located in the high back region, the Bi-low children demonstrated positional deviations from monolingual targets but the three vowels were still separated well. For the two low vowel pairs /ɛ/ and /æ/ as well as /ʌ/ and /ɑ/, the Bi-low children demonstrated substantial overlaps in F1 and/or F2. The Bi-high children, by contrast, were highly similar to monolingual English children in the front vowels but showed a certain degree of variation on the back vowels.

The Euclidean distance between the group means of selected English vowel pairs was calculated (shown in Table 6) to examine whether the two bilingual groups could separate English vowels as monolingual children did. The 10 tested English vowels were divided into four groups: high front vowels /i, ɪ, e/, mid/low

### Table 5: Summary of t-test results on trajectory length (TL) of Mandarin and English similar vowel pairs in the monolingual and bilingual children.

<table>
<thead>
<tr>
<th></th>
<th>Monolingual</th>
<th>Bi-low</th>
<th>Bi-high</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>p</td>
<td>t</td>
</tr>
<tr>
<td>M/a/-E/a/</td>
<td>2.511, 0.022</td>
<td>2.713, 0.035</td>
<td>4.097, 0.005*</td>
</tr>
<tr>
<td>M/a/-E/ʌ/</td>
<td>3.336, 0.004*</td>
<td>1.654, 0.149</td>
<td>3.316, 0.013</td>
</tr>
<tr>
<td>M/i/-E/i/</td>
<td>1.244, 0.23</td>
<td>0.457, 0.663</td>
<td>–0.617, 0.557</td>
</tr>
<tr>
<td>M/u/-E/u/</td>
<td>2.504, 0.022</td>
<td>0.634, 0.549</td>
<td>1.211, 0.265</td>
</tr>
<tr>
<td>M/y/-E/u/</td>
<td>–0.168, 0.868</td>
<td>–1.835, 0.116</td>
<td>–1.265, 0.246</td>
</tr>
<tr>
<td>M/x/-E/o/</td>
<td>3.581, 0.002*</td>
<td>3.594, 0.011</td>
<td>2.244, 0.06</td>
</tr>
</tbody>
</table>

*Represents significant results after Bonferroni adjustments for multiple comparisons.
The Euclidean distance was calculated for eight vowel pairs /i-ɪ/, /ɪ-e/, /i-e/, /ɛ-æ/, /u-ʊ/, /ʊ-o/, /u-o/, and /ʌ-ɑ/. With reference to the monolingual data, the Bi-low group showed a good separation for high front vowels but not for the vowel pairs /ɛ-æ/, /u-ʊ/, /u-o/, and /ʌ-ɑ/. The Bi-high group, by contrast, demonstrated a similar magnitude of vowel separation for most vowel pairs even though their production of back vowels showed positional deviations from the monolingual targets.

A MANOVA test was performed to examine the group difference on the normalized midpoint formant values in each language, with F1 and F2 defined as the dependent variables and group and vowel defined as the independent variables. Varying formant values as a function of vowel quality was expected and was not the focus of the present study. Therefore, no further discussion about the vowel effect was provided. For the Mandarin vowels, the analysis revealed a significant

**Table 6:** The Euclidean distance of selected English vowel pairs in the monolingual and bilingual children. The distance was calculated based on the normalized formant values.

<table>
<thead>
<tr>
<th></th>
<th>i-I</th>
<th>i-e</th>
<th>i-ɛ</th>
<th>ɛ-æ</th>
<th>u-ʊ</th>
<th>u-o</th>
<th>u-ʊ</th>
<th>ʌ-ɑ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monolingual</td>
<td>1.39</td>
<td>0.64</td>
<td>0.833</td>
<td>0.873</td>
<td>0.924</td>
<td>0.538</td>
<td>1.223</td>
<td>0.702</td>
</tr>
<tr>
<td>Bi-low</td>
<td>1.292</td>
<td>0.417</td>
<td>0.912</td>
<td>0.283</td>
<td>0.477</td>
<td>0.46</td>
<td>0.757</td>
<td>0.179</td>
</tr>
<tr>
<td>Bi-high</td>
<td>1.387</td>
<td>0.669</td>
<td>0.768</td>
<td>0.908</td>
<td>0.825</td>
<td>0.631</td>
<td>0.726</td>
<td>0.746</td>
</tr>
</tbody>
</table>
multivariate main effect of vowel (Wilks’ $\Lambda = 0.001$, $F(8,258) = 889.04$, $p < 0.0001$) but no significant multivariate main effect of group or interaction effect. For the English vowels, the analysis revealed a significant multivariate main effect of vowel (Wilks’ $\Lambda = 0.004$, $F(18,358) = 301.915$, $p < 0.0001$) and a significant interaction effect between group and vowel (Wilks’ $\Lambda = 0.662$, $F(36,358) = 2.278$, $p < 0.0001$) but no significant multivariate main effect of group. The unilateral analysis revealed that the significant interaction effect was reflected on both F1 ($F(18,180) = 2.854$, $p < 0.0001$) and F2 ($F(18,180) = 1.718$, $p = 0.04$). Subsequent analysis on individual vowels revealed that the group difference was mainly reflected in the vowels /æ/, /ʌ/, and /ɑ/. In particular, the Bi-low children were significantly different from the Bi-high and monolingual English children on F1 for the vowel /æ/. For the other two vowels, the Bi-low children were significantly different from the Bi-high children on F1 for /ʌ/ and significantly different from the monolingual English children on F2 for /ɑ/.

Figure 4 displays the group comparison of formant trajectories based on the normalized formant values for Mandarin and English, respectively. The Bi-low children showed highly compatible formant trajectories with monolingual Mandarin children for all tested Mandarin vowels except for /y/. The Bi-high children produced Mandarin /u/ with positional deviations from monolingual children and Bi-low children. In addition, the Bi-high children differed from the monolingual and Bi-low children in the general shape and direction of formant trajectories for Mandarin /y/ and /ɤ/. A two-way ANOVA was used to compare Mandarin TLs among the two bilingual groups and the monolingual children, with group and vowel defined as the two main effects. The results revealed a significant main effect of vowel ($F(4, 130) = 18.23$, $p < 0.0001$) but no group effect or group by vowel interaction effect. As for the English vowels, the two groups of bilingual children demonstrated highly compatible trajectories with monolingual English children for the three high front vowels. For the high back vowels, the trajectory of the tense vowel /u/ in both bilingual groups was different in the relative position, but similar in the general shape, direction, and magnitude features relative to monolinguals. The trajectories of /ʊ/ in both bilingual groups showed distinctive dynamic features from monolingual English children in the shape, direction, and relative position. With regard to the four vowels located in the mid/low region, both bilingual groups generally followed the monolingual English children in the direction and shape of trajectories but evident positional deviations can be observed, especially for the Bi-low children. The two-way ANOVA test on English TLs yielded a significant main effect of vowel ($F(9, 180) = 12.821$, $p < 0.0001$) and a weak group by vowel interaction effect ($F(18, 180) = 1.689$, $p = 0.045$). No group effect was found. Subsequent analysis of TL among three groups for each vowel revealed no significant difference between the bilingual and monolingual children.
This study investigated the static and dynamic acoustic features of vowel productions in 3- and 4-year-old Mandarin (L1)-English (L2) bilingual children and age-matched monolingual children. Three questions were of interest here: first, whether young bilingual Mandarin-English children successfully separate the two phonetic systems and establish vowel categories in L2 at the early stage of L2 acquisition; second, how phonetic assimilation and dissimilation are used in young bilingual children; and third, whether and to what extent young bilingual children differ from corresponding monolingual children. To answer these questions, cross-language and within-language comparisons were conducted on the midpoint formant frequency values and vowel dynamic spectral changes between the bilingual and monolingual children.

The first question addressed the development of vowel categories in L2 and the separation of L2 from L1 in young bilingual children. Previous studies pointed out that both children and adults who were immersed in an L2 early in life can produce fully distinguished phonetic categories and develop separate phonetic systems (Flege et al. 1999; Guion 2003; Lee and Iverson 2009, 2012, 2017; MacLeod and Stoel-Gammon 2005; MacLeod et al. 2009; Oh et al. 2011). While most of these group studies focused on adult bilinguals or bilingual children older than five
years of age, little phonetic data has been reported on bilingual children younger than five. Due to the lack of data from children at the very initial stage of L2 learning, the process of phonetic separation and formation remains largely unknown. In the present study, we had two groups of bilingual children (Bi-low and Bi-high) differing in the duration of immersion in an English-learning environment. The two groups of bilingual children both started to learn English around 2–3 years of age (except for two who started before six months) but the Bi-low children had a shorter duration of English immersion and were younger than the Bi-high children. Therefore, we consider the Bi-low group as representing the initial stage of English learning and the Bi-high group as representing the next developmental stage. The comparison of these two bilingual groups enables us to closely observe the development of L2 vowel categories and the separation of L2 from L1.

Of these two bilingual groups, the Bi-high children separated English vowels well and produced English vowels (both static midpoint formant values and dynamic vowel spectral change) better than the Bi-low children did. The major difference between the two bilingual groups resided in the organization of the four low vowels /ɛ/, /æ/, /ʌ/ and /ɑ/. The Bi-low children produced /ɛ/ with substantial variation in F1 and produced /æ/ at a higher position merging with /ɛ/. The vowel /ʌ/ was produced with substantial variation in F1. Some Bi-low children produced the vowel /ʌ/ at a position even lower than the vowel /ɑ/. The raised /æ/ and lowered /ʌ/ greatly changed the shape of English vowel space in the Bi-low children from a quadrilateral to a triangle, similar to the shape of Mandarin vowel space. In addition to the switch and merging of low vowels, the high back lax vowel /ʊ/ shifted up and back and the high back tense vowel /u/ moved back in the Bi-low children, which resulted in the approximation of this tense-lax pair. These observations indicate that at the very initial stage of English learning, the bilingual children’s production of the L2 vowels was driven by their native language.

Even though we claimed that the Bi-low children had not fully developed separate phonetic categories for their English, it does not mean that none of the English vowels had been acquired. Not surprisingly, the Bi-low children produced the English /e/ and /o/ vowels similar to their English monolingual peers, given that these vowels are similar to the Mandarin diphthongs /ei/ and /ou/. According to SLM, the bilingual children should have no difficulty acquiring the L2 sounds that are identical to their L1 counterparts as they could directly apply their native sounds to the new language. Furthermore, as shown in Figures 1 and 3, the Bi-low children had successfully separated the high front vowels /i/, /ɪ/, /e/. The static and dynamic features of these vowels in the Bi-low children were highly similar to those of the monolingual English children. Nevertheless, the high back vowels /u/ and /ʊ/ in the Bi-low children were not separated as well as in monolingual
children. The low vowel /ɛ/ was largely overlapped with /æ/ and /ʌ/ was merged with /a/. Yang et al. (2015) documented the developmental path of phonetic categories in one bilingual Mandarin-English child. The child separated the high vowels first and experienced a long process to separate the low back vowels. The relatively well-separated high front vowels and apparent off-target production of low front vowel /æ/ and low-mid back vowel /ʌ/ in the present study were generally compatible with the development path described in Yang et al. (2015). These observations, together, suggest that instead of separating the L2 vowel contrasts all at the same time, these young bilingual children tended to show a high to low progression (within the vowel space) in establishing their L2 vowel categories.

Referring to the phonological acquisition of monolingual English children, researchers reported a lower accuracy of /ɛ/ but higher accuracies of /u/, /o/, /i/, /ɪ/ (Davis and MacNeilage 1990), and common confusions between the low vowels /æ/, /a/ and mid-low vowel /ʌ/ (Kehoe and Stoel-Gammon 2001). While the exact acquisition order of individual vowels varied in different studies, the development path of English vowel categories in the bilingual children in the present study seems to coincide with the general trend of acquisition in monolingual children. Polka and colleagues proposed the Natural Referent Vowel (NRV) account to explain the language-universal directional asymmetries in vowel discrimination (Masapollo et al. 2017; Polka and Bohn 2011). According to this account, infants demonstrate asymmetrical discrimination in favor of peripheral vowels (e.g. /a/, /i/, /u/) over non-peripheral vowels (e.g. /ʌ/, /ɪ/, /ʊ/). As assumed in PRIMIR, both bilingual and monolingual children are equipped with the same representational spaces and acquisition mechanisms. It is reasonable to presume that this directional asymmetry bias would also be present in bilingual children. With the language input favoring more peripheral vowels, children may acquire peripheral vowels such as /a/, /i/, /u/ earlier than the non-peripheral vowels /ʌ/, /ɪ/, /ʊ/ in speech production. In the present study, the shifting of /ʌ/ to /a/ and /u/ to /u/ in the Bi-low children conforms to the NRV predictions. But the approximation of /æ/ towards /ɛ/ seems inconsistent with the NRV prediction. PRIMIR proposes that bilingual children use comparison-contrast mechanism to accommodate different task demands and keep pace with monolingual peers. Mandarin /a/ is located in the low mid region of vowel space and partially overlaps with English /æ/ and /a/. With the mid-low back vowel /ʌ/ lowered to merge with /a/, the Bi-low children might move /æ/ towards /ɛ/ to reduce the crowd of low vowels. The other plausible explanation is that the upward shift of /æ/ in the bilingual children reflects the trend of late acquisition of /æ/, as reported in the monolingual English children in previous literature (Kent 1992; Wellman et al. 1931).
Other than the directional asymmetry account, the high to low progression and late separation of low vowels /ɛ/-/æ/ and /ʌ/-/ɑ/ in our bilingual children likely reflect the experiential influence of the children’s native language. As assumed in PRIMIR and SLM, bilingual children receive different sources of language input and task demands. The different sources likely interact with each other and mutually act on the development of both languages. Of the five Mandarin monophthongal phonemes, /a/ is the only one located in the low region of vowel space. Sun and van Heuven (2007) examined the perceptual mapping of English vowels to Mandarin vowels by Mandarin adult listeners. They found that English vowels /ɛ/, æ, ã, ɔ, ɑ/ were all identified as Mandarin /a/. This perceptual data explains well that the lack of low vowel contrasts in Mandarin might cause difficulties separating English low vowels in these young children at the beginning of English learning. Another finding related to the high-to-low progression is the well-separated lax vowel /ɪ/ from tense /i/ in these bilingual children. In our previous study of Yang et al. (2015), the three-year-old child showed an early separation of lax /ʊ/ from /u/ in comparison to the late separation of low back vowels. Previous studies reported that Mandarin-speaking English learners experience considerable difficulties in distinguishing and producing English lax vowels /ɪ/ and /ʊ/ (Chen 2006; Chen et al. 2001). Although L2 studies usually compare the lax vowels /ɪ/ and /ʊ/ with the corresponding tense vowels /i/ and /u/, these two lax vowels are distant enough from the two tense vowels /i/ and /u/ in the vowel space (as shown in Figure 1). The bilingual children might treat the lax vowels /ɪ/ and /ʊ/ as new sounds that can be easier to form than similar sounds, according to SLM.

In addition to the midpoint formant values, vowel dynamic features were compared between Mandarin and English in bilingual children with reference to the monolingual targets. Some Mandarin and English vowels (e.g. Mandarin /ɤ/ and English /ʊ/) showed positional approximation in the vowel space based on midpoint values. However, the formant dynamics revealed distinctive patterns (e.g. magnitude, shape, direction of formant movement) in these two vowels. Vowel dynamic features provide important information to define the acoustic profile of vowel sounds. In the current study, the two English lax vowels /ʌ/ and /ʊ/ produced by monolingual English children had significantly shorter trajectory lengths than the two positionally similar Mandarin vowels /ɑ/ and /ɤ/ produced by monolingual Mandarin children. Of the two bilingual groups, both tended to show less formant movement for the two English lax vowels, but they failed to show monolingual-like differences in trajectory length for the tested Mandarin-English similar vowel pairs. In the meantime, the Bi-high children showed significantly different trajectory lengths between Mandarin /a/ and English /ɑ/, which was not shown in the two vowels produced by the monolingual groups. These differences
suggest that the two vowel systems in the bilingual children, even in the Bi-high children, were not separated in the same way as those in monolingual children.

The second question addressed the phonetic shift as a result of L1–L2 interactions. Previous studies reported mutual influence between L1 and L2 in adult L2 learners (Chang 2012; Flege et al. 2003; Guion 2003). Compared to adult speakers, children at a young age have not fully stabilized the phonetic system. In the meantime, children possess greater neural plasticity in language learning than adults. It is conceivable that children are more likely to show stronger and more dynamic L1–L2 interactions than adults do. Lee and Iverson (2012) reported that bilingual Korean-English children used both assimilatory and dissimilatory processes to organize the two vowel systems. Yang and Fox (2017) reported L1–L2 interactions in 5–6-year-old bilingual Mandarin-English children. Assimilatory movement of L2–L1 sounds and L1–L2 sounds were both found in these bilingual children. In the present study, the bilingual children at the very early stage of L2 learning were even younger than the participants reported in previous studies. These young bilingual children demonstrated an assimilatory movement of L2 vowels to their native vowel categories when they just started learning the new language. Later, with a longer immersion in the L2 and well-developed L2 vowel categories, the bilingual children used both mechanisms to refine the vowel organizations. Although the employment of assimilatory and dissimilatory processes varies in different bilingual populations and different studies, the underlying driving force of these adaptive mechanisms might be associated with the comparison-contrast strategy of language learning in bilingual children as proposed in PRIMIR. When children compare the two phonetic systems and notice similarities between two or more sounds, they may employ an assimilatory process to group those sounds and reinforce the similar features among them. When children detect the differences between two or more sounds, they may employ a dissimilatory process to separate them or make a contrast between them.

In the present study, Mandarin /y/, /u/ and English /u/ demonstrated a complex yet interesting interaction pattern. Although English /u/ has a phoneme counterpart of Mandarin /u/, monolingual English children produced this vowel in a more fronted position close to Mandarin /y/ (shown in Figures 1 and 2). In the two bilingual groups, both shifted English /u/ close to the Mandarin counterpart /u/ (assimilation) but away from Mandarin /y/ (dissimilation). It is noteworthy that both bilingual groups produced English /u/ with substantial variation in F2, which indicates inconsistencies in individual bilingual children. While some children might emphasize the similarity between English /u/ and Mandarin /y/ pronunciation, other children might treat English /u/ as its Mandarin counterpart /u/. Meanwhile, the Bi-high group showed greater variation in F2 for Mandarin /u/ than the Bi-low and monolingual children. This difference suggests that the two fronted
vowels Mandarin /y/ and English /u/ made some Bi-high children move Mandarin /u/ away from its target position but close to similar English sounds. The Mandarin vowel /y/ was the only one among the five monophthongs that showed observable deviations from the monolingual target in both static and dynamic features in both bilingual groups. The two bilingual groups, especially the Bi-high children, showed substantial variation in F2, which suggests that some bilingual children shifted Mandarin /y/ backward close to English /u/ (assimilation) while some children moved it away from English /u/ (dissimilation). Note that Mandarin /y/ also showed relatively large variation in F2 in monolingual Mandarin children. Previous studies reported that /y/ was the last acquired monophthongal Mandarin vowel in typically-developing children (Shi and Wen 2007). The organization of the three vowels in these bilingual children suggests the complex interference between native speech sounds and L2 sounds in the early stage of L2 acquisition.

Another noteworthy finding was the production of Mandarin /a/ and English /ʌ/ and /ɑ/ in these bilingual children. With reference to the monolingual targets, the Bi-low children lowered English /ʌ/ and shifted English /ɑ/ forward, which resulted in the approximation of English /ʌ, ɑ/ to Mandarin /a/. This reflected the assimilation of English vowels to similar native sounds. By contrast, the Bi-high children moved English /ɑ/ forward but raised English /ʌ/ up. These movements resulted in the approximation of English /a/ to Mandarin /a/ but a greater differentiation between English /ʌ/ and Mandarin /a/. While assimilation and dissimilation show opposite direction of phonetic movements between L1 and L2 sounds, the bilingual children employed both mechanisms rather than one single process to construct the two phonetic systems.

The third question focused on whether the bilingual children developed monolingual-like phonetic features for both L1 and L2 vowels. Many studies suggest that early bilinguals including both children and adults could produce L2 sounds in a monolingual-like manner while maintaining the acoustic-phonetic features of their native sounds (Darcy and Krüger 2012; Flege et al. 1999; Lee and Iverson 2012, 2017; MacLeod et al. 2009). In our results, the Bi-low children with little exposure to English preserved the vowel features in their native language very well. The Bi-high children with more exposure to English showed some observable deviations from the monolingual Mandarin children on certain vowels such as /y/ and /ɤ/. The Mandarin vowel /ɤ/ of Bi-high children was highly similar to their English /u/ in the relative position and other dynamic features. This indicates that the Bi-high children might apply the features of English vowels to similar Mandarin vowels. However, due to the lack of statistical difference on Mandarin vowel features among the three groups, the differences of native vowels in the Bi-high children with reference to monolingual Mandarin children should be interpreted cautiously. For the vowel sounds in English, both bilingual groups
showed positional deviations from monolingual English children. However, the Bi-
high children formed separate English vowel categories and approximated
monolingual English children better than the Bi-low children did.

As for the vowel dynamic features, both bilingual groups showed highly
compatible formant movement patterns with monolingual Mandarin children for
all Mandarin vowels except for /y/ (shown in Figure 4). The magnitude of formant
movement did not show significant differences between the two bilingual groups
and the monolingual children. Note that the statistical analysis revealed a sig-
nificant group by vowel interaction effect for English vowel trajectory length, but
the effect was weak and the subsequent analysis for individual vowels did not
reveal a significant difference between the bilingual and monolingual children.
Additionally, the shape and direction features of the formant trajectories in most
English vowels of the bilingual children were consistent with those of the mono-
lingual children. Therefore, while the bilingual children deviated from mono-
lingual children on the L2 features, the deviations were mainly manifested on the
relative positions of the vowel sounds.

In sum, our data suggest that young bilingual children did not establish fully
separate phonetic categories for the new language and their production of the new
sounds was affected by their native language at the very initial stage of L2 learning.
They mainly utilized an assimilatory mechanism to organize the vowel systems
and maintained the phonetic features of their native vowel sounds well. After a
certain duration, though relatively short, of intensive exposure to the L2, the
bilingual children could successfully separate the phonetic categories and
distinguish the L2 phonetic contrasts in the new language. They used both
assimilatory and dissimilatory processes to construct and refine the vowel systems
in both languages. Compared to the bilingual children at the initial stage of English
learning, the bilingual children with a longer duration of immersion in English
approximated monolingual English peers to a better extent. But when compared to
the monolingual children of each language, these experienced bilingual children
still showed observable deviations in both L1 and L2.

We acknowledge that the current findings of this paper should be generalized
with caution due to several limitations. First, there was a small number of par-
ticipants and a small size of speech samples. For future studies, a larger number of participants who are at different stages of L2 learning should be recruited to pro-
vide a more detailed developmental path of their phonetic separation. Second,
although effort had been made to control the phonetic environments in these two
languages, due to the language-specific phonotactic rules and elicitation
approach used to accommodate young children, the target vowels in these two
languages occurred in different contexts. Further, while this paper focused only on
speech production, whether or not the bilingual children showed a similar
developmental process of categorization and separation in speech perception should also be addressed. As shown in Darcy and Krüger (2012), there was an evident dissociation between production and perception in their bilingual data. It would be critical to examine when and how bilingual children distinguish the L2 phonetic contrasts in the perceptual domain. Finally, as the present study investigated only the vowel system in Mandarin-English bilinguals, more studies are needed to examine the formation of L2 consonant categories as well as phonetic-acoustic features in bilingual children from different language backgrounds.

Acknowledgment: The authors would like to thank the children and their parents for their participation.

Funding sources: None.

Author contributions: JY was responsible for the design and implementation of the research, the analysis of the results and the writing of the manuscript.

Statement of ethics: The research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki. This study has been approved by the Institutional Review Board of the University of Wisconsin-Milwaukee (IRB#:20.140). All the parents of the children who participated in this study have given their WRITTEN informed consent.

Conflict of interest statement: The authors have no conflicts of interest to declare.

References


