Chapter 12
Speech Development in Mandarin-Speaking Children

Gang Peng and Fei Chen

Abstract The unique ability to communicate through speech clearly distinguishes human beings from all other animals. Children start to produce their first words by the age of one; by four, most children have developed the ability to use their native language; by six or seven, they become veteran users of their native language. Many studies have focused on the developmental trajectories of one or two types of phonetic units of speech (e.g., consonants, vowels, or tones), but a more comprehensive picture of speech development is still lacking. Questions deserving further investigation include: What are the order and rate of acquisition of various phonetic units? What is the possible driving force underlying the developmental order? How can we ascertain when children have obtained the same speech competence as adults? In this chapter, we will first review the literature on the development of speech perception and production in Mandarin-speaking children. Then, we will discuss relevant issues, and suggest possible solutions to the unresolved questions.

12.1 Introduction

Children's language acquisition provides an opportunity for us to observe a language in its nascent state and to trace it through the many subsequent changes. In learning to communicate, children need to gain knowledge of the phonological forms of their mother tongue, and gradually acquire the perceptual discrimination and articulatory gestures required to perceive and produce these sounds in an adult-like manner. For more than 100 years (dating from Sterns' diaries describing language use in infancy, 1907), there have been thousands of descriptive and experimental studies on children's speech development in different language backgrounds. Based on these findings, we can figure out the normative pathways to speech development—the

G. Peng (✉) · F. Chen
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China
e-mail: gpeng@polyu.edu.hk

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*,
Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-15-7606-5_12

age of acquisition of speech sounds (phonological acquisition)—and subsequently discuss fundamental theories of spoken language development.

To date, research in the field of child language development has focused primarily on children's acquisition of Romance and Germanic languages. Not surprisingly, English has received the most attention. Some norms of the phonological acquisition of English-speaking children, including the developmental stage of phonemes and error patterns, have been studied extensively and well established (e.g., Dodd et al., 2003; Prather et al., 1975; Stoel-Gammon & Dunn, 1985). To compare the developmental patterns among children acquiring different languages, there was a remarkable upsurge of interest in cross-linguistic studies of language acquisition (Slobin, 1985). Many studies have examined developmental universals and language-specific patterns in the developmental patterns of children from various language backgrounds by investigating the order and rate of acquisition of phonemes. Among them, the peculiarities of Mandarin Chinese offer an excellent, perhaps unique, opportunity for the evaluation and expansion of language theories and language acquisition. Nevertheless, the field of Mandarin acquisition (i.e., studies of how native Mandarin-speaking children acquire their native language) remains relatively underexplored. Furthermore, in the existing research concerning Mandarin morphology, syntax, and writing system acquisition, the acquisition of Mandarin phonology is perhaps the least explored (Hua, 2002).

Modern Mandarin is a tonal language with a relatively simple syllable structure (see Fig. 12.1). Each syllable must be attached to one of the four lexical tones which carry different lexical meanings (Wang, 1973). The expression of tones is superimposed on other phonetic units of speech, such as vowels and consonants. Tones are typically instantiated on the vowels, but their realization interacts with surrounding consonants (Hombert, Ohala, & Ewan, 1979). The lexical tones and nuclear vowels are compulsory elements of Mandarin syllables, whereas onset and ending consonants are optional (Wang, 1973). It is expected that the developmental patterns of Mandarin-speaking children's phonological acquisition reflect both universal tendencies and language-specific constraints.

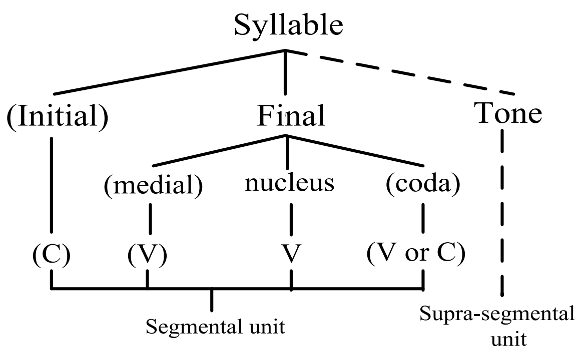

Fig. 12.1 Diagram of the modern Mandarin syllable (elements in parentheses are optional) (adapted from Chen, Peng, Yan, & Wang, 2017)

Chao (1951) presented the earliest description of the phonological acquisition of Mandarin-speaking children. This study provided an analysis of the consonant, vowel, and tone repertoires of the phonological system of a twenty-eight-month-old girl who was acquiring Mandarin as her first language. Since then, many research papers and books have focused on the acquisition of speech sounds in Mandarin. The limited information available suggests that Mandarin children's phonological development is influenced by the characteristics of Mandarin speech. This chapter will first summarize previous findings on the phonological acquisition of Mandarin, including the developmental trajectories of individual phonetic units of speech (e.g., consonants, vowels, and tones) before discussing theories of phonological acquisition to account for the detailed evidence presented in Mandarin studies. Following this, we will discuss the method of determining when children obtain the same speech competence as adults during the process of phonological maturation. In the final section, we will focus on unsolved issues, and propose future research directions that could potentially inform and advance the field of Mandarin speech acquisition.

12.2 Developmental Trajectories of Phonetic Units of Speech

Previous behavioral studies, both cross-sectional and longitudinal, potentially provide 'normative data' on the phonological acquisition by Mandarin-speaking children, which can be used for cross-linguistic comparison and the assessment of phonological disorders in Mandarin-speaking children (Dodd et al., 2003). However, significant discrepancies have also been reported in the order and age of acquisition of Mandarin phonetic units. Conflicting results in the sequence and timeline by which children master the speech sounds can be attributed to many factors, including but not limited to, criteria used for performance evaluation, collection approach (cross-sectional or longitudinal research), speech mode (spontaneous production or imitation), sample size, age range of subjects, living area of subjects, and number of transcribers.

Most concerns are related to methodological issues, particularly the criteria used. Importantly, the acquisition of speech sounds occurs gradually and progressively. It is not all or nothing (Olmsted, 1971). Since the acquisition of phonological production is a developmental continuum ranging from an initial stage of being able to articulate a proper sound of a certain phoneme to the final stage of being able to articulate that phoneme systematically accurately, the distinction between 'phonological emergence' and 'phonological stabilization' is crucial (Hua, 2002). According to Hua (2002), 'phonological emergence' determines when a child is able to articulate a certain phoneme. A phoneme is considered to have emerged when children of a certain age first produce a sound of it correctly at least once. Phonological stabilization refers to when a child articulates a phoneme with certain consistency. Since there is a certain amount of inconsistency in children's production, a criterion is needed to determine the age of phonological stabilization. A phoneme is considered stable when the child produces it correctly on at least two out of three attempts. When 90% of the children in an age group achieve an accuracy rating of at least 66.7% (i.e., 2/3) for a phoneme, it is considered to be stabilized in that age group.

12.2.1 The Acquisition of Mandarin Consonants

There are 22 consonants in Mandarin, of which 21 serve as syllable-initial consonants, and two as syllable-final consonants [ŋ, −n]. The consonant [n] can appear in both the syllable-initial and syllable-final positions, whereas the consonant [ŋ] occurs only in the coda position. Aspiration, a distinctive feature in Mandarin, is used to differentiate six minimal pairs: [p/pʰ, t/tʰ, k/kʰ, ts/tsʰ, tʂ/tʂʰ, tɕ/tɕʰ]. The places and manners of articulation of Mandarin consonants are listed in Table 12.1.

Similar to English, consonant acquisition in Mandarin-speaking children has received more attention than other types of phonetic unit. Table 12.2 shows 13 well-cited studies on Mandarin consonant acquisition. Not surprisingly, the differences among them, especially in the criteria used, resulted in differences in the age of acquisition of Mandarin consonants. Specific findings of each study are summarized as follows.

Chao (1951) provided a detailed description of the consonant acquisition of a girl who was acquiring Mandarin as her first language in the USA. At the age of 2;4, the child's consonant inventory consisted of 11 phonemes, including three pairs of unaspirated and aspirated voiceless plosives [p, pʰ, t, tʰ, k, kʰ], two nasals [m, n], and three fricatives [f, s, x]. Given that it was based on one child only, few generalizations can be made from Chao's study.

Though Jeng (1979) used the terms 'emergence' and 'stabilization,' he did not define specific criteria. This study is an attempt to test the applicability of Jakobson's

Table 12.1 Consonants in Mandarin

Manners places	Stop		Nasal	Fricative		Affricate		Lateral
	Unasp.	Asp.		Vls	Vd	Unasp.	Asp.	
Labial	p	pʰ	m					
Labio-dental				f				
Alveolar	t	tʰ	n	s		ts	tsʰ	l
Retroflex				ʂ	ʐ	tʂ	tʂʰ	
Alveolo-palatal				ɕ		tɕ	tɕʰ	
Velar	k	kʰ	ŋ	x				

Note Unasp. = Unaspirated, Asp. = Aspirated, Vls = Voiceless, Vd = Voiced

Table 12.2 An overview of studies on Mandarin consonant acquisition in chronological order

Studies	Criteria used	Collection approach	Speech mode	Sample size	Age range	Living area
Chao (1951)	Consonant emergence	Description at a certain age	Spontaneous production	1	2;4	USA
Jeng (1979)	Not clear	Longitudinal research	Spontaneous production	2	0;2–1;8 and 1;3–2;7	Taiwan
Wu and Xu (1979)	Consonant emergence	Longitudinal research	Spontaneous production	5	Birth–3;0	Not clear
Hsu (1987)	Not clear	Cross-sectional research	Spontaneous production and imitation	28	1;0–6;0	Taiwan milieu
Shiu (1990)	Consonant stabilization	Longitudinal research	Spontaneous production	2	0;9–3;0 and 0;7–2;4	Taiwan
Hua and Dodd (2000)	Consonant emergence and stabilization (66.7% criteria)	Cross-sectional research	Spontaneous production and imitation	129	1;6–4;6	Beijing
Hua (2002): Chap. 4	Consonant emergence and stabilization (66.7% criteria)	Longitudinal research	Spontaneous production and imitation	4	1;1.15–2;0.15; 1;0.0–2;0.15; 0;10.15–2;0.15; 1;2.0–1;8.0	Beijing
Cao (2003)	Consonant emergence	Longitudinal research	Spontaneous production	1	Birth–1;9	Shandong province
Si (2006)	Consonant emergence and stabilization	Longitudinal research	Spontaneous production	1	2;0–5;0	Beijing
Liu (2007)	Not clear	Longitudinal research	Spontaneous production	10	(1;6 ~ 1;11)–(1;10 ~ 2;3)	Shanghai

(continued)

Table 12.2 (continued)

Studies	Criteria used	Collection approach	Speech mode	Sample size	Age range	Living area
Xie (2009)	Not clear	Cross-sectional research	Spontaneous production and imitation	149	2;4–6;0	Shijiazhuang (Hebei Province)
Chen and Kent (2010)	Consonant emergence	Cross-sectional research	Spontaneous production	24	0;7–1;0 and 1;1–1;6	Taiwan
Zhong (2013)	Consonant emergence and stabilization (90% criteria)	Longitudinal research	Spontaneous production and imitation	1	3;4–3;10	Tianjin

Table 12.3 Age of emergence and stabilization of syllable-initial consonants as found by Hua and Dodd (2000)

Age	Consonant emergence	Consonant stabilization
1;6–2;0	t, t^{h}, k, m, n, x, tɕ, tɕ^{h}, ɕ	t, m
2;1–2;6	f, s, tʂ	n
2;7–3;0	p, l	p, t^{h}, f, x, ɕ
3;1–3;6	p^{h}, k^{h}, tʂ^{h}	k, k^{h}
3;7–4;0	ʂ	p^{h}
4;1–4;6	ts, ts^{h}, ʐ	l, s, ʑ, tɕ, tɕ^{h}
> 4;6		ʂ, tʂ, tʂ^{h}, ts, ts^{h}

laws of irreversible solidarity[1] in the acquisition of Mandarin phonology. Two boys aged 0;2–1;8 and 1;3–2;7, who were acquiring Mandarin in Taiwan, provided the speech data. Jeng found that the consonants acquired earliest were [p, t, k, ts], followed by nasals, aspirated stops, fricatives (except [f]), approximant [ʐ], and finally [f].

Wu and Xu (1979) longitudinally analyzed speech sounds of five children over three years from birth. The results indicated that the first two consonants to emerge were glottal [h] and nasal [m], before three months. Subsequently, the stops [p], [t], [k], the nasal [n], and fricative [f] emerged after three months.

Hsu (1987) conducted a small-scale cross-sectional study to examine the phonological development of 28 children aged 1;0–6;0 who were acquiring Mandarin Chinese in Taiwan. Two-thirds of the parents spoke Taiwanese as their first language, and the others spoke various Chinese dialects. Among the consonants acquired before 20 months were [p, m, t, k], while the children aged 4;4–6;0 still made errors with the consonants [f, n, tʂ, tʂ^{h}, ʂ, ʑ, ts, ts^{h}].

Shiu (1990) analyzed the phonological development of a boy aged 1;0–3;0 and a girl aged 0;7–2;4, and used phonetic accuracy and consistency in the children's realizations as explicit criteria for acquisition. The early acquisition of [p] and [m] was found to be followed by the establishment of the labial-dental contrasts between [p] and [t], and [m] and [n].

Table 12.3 summarizes the age of emergence and stabilization of Mandarin syllable-initial consonants found in Hua and Dodd's cross-sectional study (Hua & Dodd, 2000).

By 4;6, 90% of the children were able to articulate all of the 21 syllable-initial consonants. Among the first sounds produced by 90% of the children were nasals, alveolar stops, alveolo-palatal fricatives and affricates, and velar stops and fricative. The velar fricative [x] and three alveolo-palatals ([tɕ, tɕ^{h}, ɕ]) emerged very early in the children's speech. Two alveolar affricates ([ts, ts^{h}]) and the retroflex [ʐ] appeared last.

[1] The laws of irreversible solidarity were systematically proposed by Jakobson (1941/1968), and this theory made a connection between the children's phonological acquisition order and the distribution of phonological features among the world's languages. See more details in Sect. 12.3 in this chapter.

The stabilization of syllable-initial consonants in Mandarin phonology can be classified into the following three groups: (1) phonemes which were stabilized as soon as the children were able to articulate them (e.g., [t, m, p]); (2) phonemes which took a relatively short period to stabilize after the children were able to articulate them (e.g., [n, f, x]); (3) and phonemes which took a long time to stabilize after the children were able to articulate them (e.g., [tɕ, tɕ^h, s]).

Hua (2002) longitudinally examined the patterns of acquisition of Mandarin consonants among four subjects, with a particular focus on those consonants acquired before the age of two. By the end of the data collection period (24 months old), syllable-initial consonants [p, t, m] and syllable-final consonants [-n, ŋ] stabilized in the speech of all the children. There were variations in the emergence of sounds: while two subjects produced all the Mandarin consonants, one never used the sounds [p^h, k^h, ʂ, tɕ^h, tʂ^h, ts^h]. Another subject never produced the sounds [p^h, t^h, k^h, ʑ, tɕ^h, tʂ^h, ts^h] in his speech by the offset time of his data collection (i.e., 1;8).

Cao (2003) depicted her own daughter's phonological development from birth to 21 months. She found that the development of Mandarin consonants was closely related to the place and manner of consonant articulation. The developmental order was from front and back-to-middle consonants. The front consonants [m, p] and back consonant [ŋ] emerged earliest, while [ʐ] and [l] were the last to appear. Furthermore, the manner of articulation also influenced the emergence of consonants: nasals appeared first, then stops, fricatives, affricates, and lastly laterals. Voiceless consonants were realized initially as the corresponding voiced ones. The unaspirated consonants appeared earlier than the aspirated ones, and the aspirated consonants developed from weaker to stronger aspiration.

Si (2006) made a detailed observation of her daughter's phonological acquisition from two to five years of age. The age of emergence and stabilization of all Mandarin consonants found by Si are shown in Table 12.4. Unaspirated sounds were acquired earlier than the corresponding aspirated ones. Her daughter acquired all the stops by three years of age; the alveolo-palatals by 3;6; and nasals and lateral by four years old. Among the six fricatives, [ɕ] was the first to be acquired, earlier than [x], both of which were acquired before 3;6. Among the six affricates, the first acquired sound was the unaspirated alveolo-palatal affricate [tɕ], which emerged almost at the same

Table 12.4 Age of emergence and stabilization of Mandarin consonants found by Si (2006)

Age	Consonant emergence	Consonant stabilization
2;0	p, m, t, t^h, n, ŋ, f, tɕ, ɕ, l	m, t
2;6	x, s, tɕ^h, −n, p^h, k, k^h, ts	p, p^h, t^h, ɕ
3;0	ts^h, tʂ, tʂ^h, ʂ, ʐ	tɕ, k, k^h
3;6		tɕ^h, f, x
4;0		n, ŋ
4;6		l, −n
5;0		s, ʂ, ts, ts^h, tʂ, tʂ^h, ʐ

time as the velar stops [k, k^h]. The labial [m] was the earliest of the three nasals to be acquired, at approximately age two.

In Liu (2007), the front consonants were acquired relatively earlier than the back, and unaspirated consonants were acquired earlier than the aspirated ones. The stops and affricates were acquired earlier than the fricatives, and the nasals were earlier than laterals. Specifically, the nasal coda [-n] was acquired later than the initial [n].

Xie (2009) investigated the acquisition order of syllable-initial consonants in Mandarin, based on a cross-sectional study of 149 children. Front stops were found to be acquired earlier than the back ones, while there was no such regularity for fricatives and affricates. Unaspirated segments were acquired before their aspirated counterparts, and the feature of aspiration was acquired between 2;6 and 3;0. Stops were generally acquired earlier than fricatives, and most fricatives were acquired earlier than affricates, with the exception of [s]. Among the fricatives, [x] and [ɕ] were acquired earliest, followed by [f]. Regarding the place of articulation, the alveolo-palatal fricative and affricates were acquired earlier than their retroflex counterparts, and the retroflex fricative and affricates were acquired earlier than their alveolar counterparts.

In Chen and Kent (2010), the early development of consonantal production in infants learning Mandarin was studied during the transition from babbling (0;7–1;0) to producing first words (1;1–1;6). Consonantal development showed two universal patterns: labials and alveolars (including alveolars, retroflexes, alveolo-palatals) occurred more frequently than velars; and nasals developed earlier than fricatives, affricates, and liquids. They also found two language-specific patterns in Mandarin: alveolars were more prominent than labials, and affricates were developed early.

Zhong (2013) investigated the syllable-initial consonant acquisition of Mandarin-speaking children aged between 3;4 and 3;10. During this period, stops were acquired in the order [p] > [p^h] > [t] > [k] > [t^h] > [k^h] (>means 'is acquired before'); fricatives were acquired in the order [ɕ] > [f] > [x] > [s] > [ʂ]; affricates were acquired in the order [tɕ] > [$tɕ^h$] > [ts] = [tʂ] > [ts^h] = [$tʂ^h$]; and Mandarin sonorants were acquired in the order [m] > [l] > [n] > [ʐ].

In conclusion, a comparison of these studies reveals the following significant anomalies. First, they differ greatly in terms of the criteria for defining when a sound would be considered acquired ('emerged' or 'stabilized'), which led to differences in the age of acquisition identified. Some studies did not even offer a straightforward criterion (e.g., Jeng, 1979; Hsu, 1987; Liu, 2007; Xie, 2009). Reasonably, phonological emergence was much earlier than phonological stabilization (see a direct comparison within the same studies in Tables 12.3 and 12.4).

Second, larger samples of subjects were often recruited in the cross-sectional studies than in the longitudinal ones. Consonants often emerged earlier in the longitudinal studies than in the cross-sectional study. For example, all Mandarin consonants emerged before three years in the longitudinal study by Si (2006), but was delayed to 4;6 in the cross-sectional study by Hua and Dodd (2000). The study by Dodd (1995) also reported that the English phoneme repertoires found in the longitudinal studies consisted of more phonemes than those found in the cross-sectional study. A plausible explanation for this difference may be that different types of speech sample were collected. Although both types of study attempted to collect spontaneous speech samples, the children in the longitudinal study had the opportunity to partially control topics and content, and to produce familiar words during daily communication with caregivers, and were less stressful and nervous than in the picture-naming task which is often adopted in cross-sectional study. Since some longitudinal studies involved parents (often also the authors) transcribing their own children's speech samples (e.g., Cao, 2003; Si, 2006; Zhong, 2013), another possibility for earlier emergence in the longitudinal studies could be familiar listeners giving credit for intended phonemes, even if these phonemes were imperfectly realized.

Third, in those studies which adopted the picture-naming task, some of the younger children failed to produce the target word spontaneously and were then asked to imitate the examiner (e.g., Hsu, 1987; Hua & Dodd, 2000; Xie, 2009; Zhong, 2013). Since previous studies quantified the gain in speech intelligibility with the help of 'lip reading' in comparison with the acoustic signal only (e.g., Benoît and Goff, 1998), children in these studies might have learnt the speech production from the examiner at the testing time, which might not reflect their actual sound production ability.

Fourth, although children in all these studies were acquiring Mandarin as their first language, their language environments were not the same, as summarized in the last column of Table 12.2. For example, the children in Chao (1951) and Clumeck's (1980) studies were acquiring Mandarin in America, but those in Jeng (1979), Hsu (1987), and Shiu's (1990) studies were in Taiwan, influenced by Taiwanese (or Hokkien, of the Southern Min dialectal group of Chinese) on a daily basis. It is not clear how much influence the L2 language environment (another language or another Chinese dialect) may exert on the age and order of Mandarin phonemes' acquisition.

Finally, the existence of individual variations in phonological development may be a cause for the disagreements in the above studies, as many longitudinal findings were obtained with data from only one case study (e.g., Chao, 1951; Cao, 2003; Si, 2006; Zhong, 2013).

Despite the differences in the analyses and findings mentioned above, these studies have reached a great deal of consensus on the order and rate of acquisition of some Mandarin consonants. First, consonants with different manners were acquired in a specific order, with stops and nasals appearing earlier than fricatives and affricates, which were earlier than laterals. The voiceless consonants were always realized before the corresponding voiced ones in young children. Similarly, unaspirated consonants were acquired earlier than the corresponding aspirated ones in most cases. Second, among the six Mandarin fricatives, [ɕ] and [x] were always the first ones to be acquired, while [s] was often acquired later. The earlier acquired affricates were always the alveolo-palatal ones [tɕ] and [tɕʰ], while the other four affricates were often the last ones to be acquired among all 22 Mandarin consonants. Third, the labial consonants [m] and [p] and the alveolar [t] were usually among the earliest Mandarin consonants acquired. The lateral sound [l] and four retroflex sounds [ʂ, ʐ, tʂ, tʂʰ] were always among the last Mandarin consonants acquired.

12.2.2 The Acquisition of Mandarin Vowels

Although analyses of the Mandarin vowel system remain somewhat unsettled on the number of surface vowels and the distribution of allophones in the phonemic system, most studies proposed 12 or 13 surface vowels and four to six vowel phonemes (Cheng, 1973; Lin, 1989). Table 12.5 lists all Mandarin monophthongs, diphthongs, and triphthongs. Among the Mandarin linguofacial monophthongs, [ɛ] and [ə] both occur in very restricted contexts: [ɛ], as a monophthong, is used only in conversational particles expressing a speaker's emotions, such as surprise and agreement; [ə], as a monophthong, occurs only in weakly stressed syllables. The vowel [ɚ] is a retroflexed central vowel, which occurs either in isolation or retroflection and thus has a very restricted combination with onset consonants. The apical vowel [ɿ] occurs only after [ts], [ts^h], and [s], while the apical vowel [ʅ] occurs only after initials [tʂ], [$tʂ^h$], [ʂ], and [ʐ]. There are nine diphthongs and four triphthongs collectively.

Table 12.6 lists 11 well-cited studies on Mandarin vowel acquisition, with details including criteria, collection approach, speech mode, sample size, age range, living area of subjects. Similar to Mandarin consonant acquisition, differences in these areas, especially the criteria used, lead to different ages of acquisition of Mandarin vowels being identified. Specific findings of each study are summarized in the following.

In Chao (1951), among the observed patterns of Mandarin vowel acquisition, diphthongs tended to be realized as monophthongs, the sounds [i, u, y] in diphthongs and triphthongs went through stages of deletion and addition before stabilization.

Wu and Xu (1979) found that the early occurring Mandarin vowels produced by infants were [a], [ɛ], [i], which belong to the category of unrounded front vowels.

In Jeng (1979), the four vowels [A, ɑu, i, ɛ] occurred earliest, while [u, y, o] appeared later.

Hsu (1987) found that the monophthongs [A] and [i] emerged first, around the age of 1;1. By the age of 1;6, all monophthongs had emerged, with the exception of

Table 12.5 Vowels in Mandarin

	Vowel height	Vowel backness		
		Front	Central	Back
Monophthong (linguofacial vowel)	High	[i] [y]		[u]
	High-mid		[ə]	[ɤ] [o]
	Low-mid	[ɛ]		
	Low		[A]	
Monophthong (apical vowel)	[ɿ], [ʅ]			
Retroflex vowel	[ɚ]			
Diphthong	[iA], [uA], [uo], [iɛ], [yɛ], [ai], [ei], [ɑu], [ou]			
Triphthong	[uai], [uei], [iɑu], [iou]			

Table 12.6 An overview of studies on Mandarin vowel acquisition, in chronological order

Studies	Criteria used	Collection approach	Speech mode	Sample size	Age range	Living area
Chao (1951)	Vowel emergence	Description at a certain age	Spontaneous production	1	2;4	USA
Wu and Xu (1979)	Vowel emergence	Longitudinal research	Spontaneous production	5	Birth–3;0	Not clear
Jeng (1979)	Not clear	Longitudinal research	Spontaneous production	2	0;2–1;8 and 1;3–2;7	Taiwan
Hsu (1987)	Not clear	cross-sectional research	Spontaneous production and imitation	28	1;0–6;0	Taiwan milieu
Hua and Dodd (2000)	Vowel emergence and stabilization (66.7% criteria)	Cross-sectional research	Spontaneous production and imitation	129	1;6–4;6	Beijing
Hua (2002): Chap. 4	Vowel emergence	Longitudinal research	Spontaneous production and imitation	4	1;1.15–2;0.15; 1;0.0–2;0.15; 0;10.15–2;0.15; 1;2.0–1;8.0	Beijing
Cao (2003)	Vowel emergence	Longitudinal research	Spontaneous production	1	Birth–1;9	Shandong province
Si (2006)	Not clear	Longitudinal research	Spontaneous production	1	2;0–5;0	Beijing
Liu (2007)	Not clear	Longitudinal research	Spontaneous production	10	(1;6–1;11)–(1;10–2;3)	Shanghai
Shi and Wen (2007)	Vowel stabilization (66.7% criteria)	Cross-sectional research	Spontaneous production and imitation	40	1;0–6;0	Tianjin
Chen and Kent (2010)	Vowel emergence	Cross-sectional research	Spontaneous production	24	0;7–1;0 and 1;1 to 1;6	Taiwan

[y], which presented difficulties even to children aged 6;0. The diphthongs emerged almost as early as the monophthongs, but only five diphthongs [ai, ɑu, iA, iɛ, uA] stabilized by the age of 6;0.[uai] stabilized first, at the age of 1;9, followed by [iɑu] between 2;7 and 3;0.[uei] and [iou] did not stabilized even between the ages of 5;1 and 6;0.

In Hua and Dodd (2000), vowels were found to emerge very early in development. The youngest group of children (1;6–2;0) was able to produce all the monophthongs.

Diphthongs were often reduced to monophthongs. Triphthongs were often reduced to diphthongs (in most cases) or sometimes to monophthongs.

In Hua (2002), among the monophthongs, the central low vowel [A] and back high vowel [u] were the earliest to emerge in the four children studied here; the retroflex vowel [ɚ] and the back vowel [o] seemed to be the last monophthongs to emerge in the children's output.[ei] was the first diphthong to emerge for all children, and [yɛ] the last.[iou] was the first triphthong to emerge, while [uai] was the last, for three children.

In Cao (2003), the emergence order of Mandarin monophthongs was identified as [A] > [i] > [ɛ] > [o] > [ɤ] > [ɚ] > [u] > [y]. The Mandarin finals with diphthongs and triphthongs emerged later than monophthongs. Nasal finals were the last to appear, with the nasal coda emerging and developing gradually.

In Si (2006), the linguofacial vowel occurred earliest, followed by the retroflex vowel. The two apical vowels appeared later. Among the linguofacial vowels, monophthongs were acquired at age two, one year earlier than the diphthongs and triphthongs. The retroflex vowel [ɚ] was acquired at age 3;6, while the apical vowels [ɿ] and [ʅ] were not fully acquired even at age five.

In Liu (2007), for Mandarin monophthongs, the acquisition order was as follows: linguofacial vowels > back apical vowel [ʅ] > retroflex vowel [ɚ] > front apical vowel [ɿ]. Diphthongs were acquired earlier than triphthongs.

Shi and Wen (2007) conducted a cross-sectional study of Mandarin vowels produced by 40 Mandarin-speaking children aged one to six. Their results indicated that the acquisition order (66.7% criteria) of Mandarin monophthongs was as follows: [A] > [i] > [ɤ] > [u] > [ʅ] > [ɿ] > [y]. The developmental phases of Mandarin vowels could be divided into three stages: before age two, from two to three, and after three. The development of vowels [A], [i], [ɤ] occurred before age two, and there was tremendous progress in [u], [ʅ], [ɿ], and [y] between two and three. After three, the development speed of all Mandarin vowels slowed down and became comparatively stable.

Chen and Kent (2010) recoded spontaneous vocalizations produced by 24 infants grouped by age: G1 (0;7–1;0) and G2 (1;1–1;6). Vowel development exhibited two universal patterns: the predominance of low and mid vowels, e.g., [ɛ] and [ə], over high vowels. Language-specific patterns were also found, such as the early appearance and acquisition of low vowels [A]. Vowel production was similar in G1 and G2, and a continuum of developmental changes brought infants' vocalization closer to the adult model.

In conclusion, due to methodological differences among various studies, the significant discrepancies shown in consonant acquisition also applied to the age and order of Mandarin vowel acquisition. However, these studies have reached a great deal of consensus in terms of Mandarin vowel acquisition. First, the Mandarin diphthongs and triphthongs were acquired later than monophthongs. Diphthongs and triphthongs tended to be realized as monophthongs during the process of vowel acquisition. Second, among the monophthongs, the linguofacial vowels were acquired first (with the exception of [y]), while the retroflex vowel [ɚ] and two apical vowels ([ɿ], [ʅ]) appeared later. Third, the unrounded Mandarin vowels were often acquired earlier than the rounded vowels ([u], [y]). The low vowel [A] and high vowel [i] were always among the early acquired Mandarin vowels.

12.2.3 The Acquisition of Mandarin Tones

As mentioned, Mandarin is a tonal language that exploits variations in pitch at the syllable level to distinguish lexical meanings. The four lexical tones can be categorized phonologically into a high-level tone (Tone 1), a mid-rising tone (Tone 2), a low-falling-rising tone (Tone 3), and a high-falling tone (Tone 4). The major factors distinguishing different lexical tones are the height and direction of the fundamental frequency (F0) contour. The F0 contour of Tone 3 varies depending on context. It is typically a dipping (low-falling-rising) tone in isolation and a low falling tone in non-final position (Xu, 1997). Moreover, when Tone 3 is produced in continuous speech and followed by another tone 3, it changes into Tone 2. There are some other tone sandhi rules in Mandarin, which are closely associated with the morphological structures of Chinese words, and sometimes with grammatical structures. Moreover, weak stress, often referred to as the 'neutral tone' or weak syllable (see Norman, 1988), is one of the essential prosodic features in Mandarin. This chapter only focuses on the acquisition of four Mandarin citation tones (Tones 1–4) occurring at the monosyllabic level.

The findings regarding Mandarin tone acquisition are comparatively more consistent than those for the acquisition of Mandarin vowels and consonants. Studies on tone production suggested very early mastery of Mandarin tones and reported that children produce tones correctly around age two, well before they have achieved mastery of consonants and vowels (e.g., Chao, 1951; Clumeck, 1980; Hua & Dodd, 2000; Hua, 2002; Li & Thompson, 1977; Si, 2006). Furthermore, Tone 1 and Tone 4 have been found to be mastered earlier than Tone 2 and Tone 3 (Clumeck, 1980; Hua, 2002; Li & Thompson, 1977), and most of the tone errors involve a lack of distinction between Tone 2 and Tone 3 (e.g., Clumeck, 1977; Li & Thompson, 1977).

Yeung, Chen, and Werker (2013) systematically explored the developmental changes in tone perception in Mandarin-speaking infants (see also Tsao & Liu, Chap. 10 this volume). Their results demonstrated that language experience could affect the perception of lexical tones from as early as four months old: English-, Cantonese-, and Mandarin-exposed infants demonstrated different discrimination abilities that accorded with the properties of their native languages at this stage. This study suggested that the formation of tone categories took place earlier than that of vowels and consonants. What was previously regarded as a language-general stage of phonological development (from birth to six months) appears not to be applicable to infants whose mother tongue is a tonal language. Moreover, Tsao (2008) examined whether the acoustic similarity between lexical tones would affect the perceptual discrimination performance of ten- to 12-month-old infants using the head-turn methodology. Infants were taught to turn their heads to a sound or a change in sound sequences. The results showed that the discrimination accuracy between Tone 1 and

Tone 3, which is the most distinct contrast acoustically, was greater than that for other less distinct tonal contrasts (e.g., Tone 2 vs. Tone 3). In a picture-pointing task, three-year-old children showed greater accuracy in perceiving Tone 1, Tone 2, and Tone 4 (90%, 87%, and 89%, respectively) than Tone 3 (70%), which was most frequently misidentified as Tone 2 (Wong et al., 2005).

In conclusion, it seems that Mandarin tones are fully acquired earlier than segmental elements (vowels and consonants). The production and perception of Mandarin Tone 1 and Tone 4 are acquired earlier than Tone 2. Tone 3 is always the last tonal category to be acquired in Mandarin. The patterns of tone acquisition identified here raise two specific questions: Why is tone acquisition completed earlier than that of segments? Why is one tone acquired earlier than another? We will discuss these questions in the next section along with related theories of phonological acquisition.

12.3 Theoretical Approaches to Mandarin Phonological Acquisition

The theory of child phonology began with Jakobson's (1941/1968) monograph *Child language, aphasia, and phonological universals* (translated from German in 1968), which is probably the best-known and most influential account of phonological development. It is grounded in the framework of structural linguistics. This theory suggests that the early acquisition of a sound depends on its distribution across the world's languages.

The acquisition of Mandarin phonetic units partially supports Jakobson's theory. According to his 'laws of irreversible solidarity,' nasals should be acquired before orals, front consonants before back consonants, and stops before fricatives. Shiu (1990) found that the early acquisition of [p] and [m] was followed by the establishment of the labial-dental contrasts between [p] and [t], and [m] and [n]. This finding agrees with Jakobson's postulation on the first and second consonant split (i.e., the first contrast within the consonantal system is between nasal and oral; the second between labial and dental). However, according to Hua and Dodd (2000), front consonants ([f]) are acquired at about the same stage as back consonants ([x, ŋ]). Moreover, the alveolo-palatal consonants [tɕ, $tɕ^h$, ɕ] tend to be acquired earlier than their alveolar counterparts [ts, ts^h, s] which have a more frontal articulation. These findings clearly contradict Jakobson's predictions of the earlier acquisition of front consonants. Moreover, a sound or feature with high distribution frequency in the world's languages should be acquired early (Jakobson, 1941/1968), and vice versa. The late acquisition of Mandarin vowel [y], which is a rather rare sound among other languages, greatly supports this notion. Furthermore, high falling tones are more frequent across the world's languages than rising tones, and one study (Chen & Kent, 2009) indicates that in Mandarin-learning infants, falling contours occur (Tone 4) significantly more often than rising contours (Tone 2), similar to the prosodic patterns found in English-learning infants during the first year of their lives. However, the three alveolo-palatal affricates ([tɕ], [tɕʰ], [ɕ]), which are very rare in the world's major languages, emerge very early in Mandarin. These data do not support Jakobson's proposal that the frequency of a phoneme across the world's languages reflects its age of acquisition.

The notion of 'markedness' has also been used to interpret similarities and differences in the order of sound acquisition (Eckman, 1977). It has been hypothesized that those sounds appearing early in a child's phonological inventory are maximally unmarked, while those occurring late are marked. Therefore, children use unmarked sounds as substitutions for marked sounds in early stages. The unmarked features are assumed to acquire first because they are considered more phonetically natural, and the marked features (e.g., aspiration, roundedness, retroflex, and falling-rising pitch direction) are acquired later. Therefore, children tend to replace marked features with unmarked features.

Jakobson's 'laws of irreversible solidarity' and his theory of 'markedness' sought to explain children's acquisition of sounds in relation to the structures of the languages they are learning. In contrast, other researchers (Kent, 1992; Locke, 1983) emphasized the role of child-centered articulatory and perceptual constraints on children's acquisition of phonology. The phonemes acquired later in Mandarin include all retroflex sounds, liquids, and rounded vowels. The late acquisition of these sounds, which are believed to be difficult to articulate and perceive (Locke, 1983), supports the hypothesis that biological constraints affect the order of phonological acquisition. Furthermore, in terms of the acquisition order of Mandarin lexical tones, Wong (2012) indicated that the order of accuracy of Mandarin children's four tones (i.e., Tone 4, Tone 1, Tone 2, and Tone 3, from highest to lowest accuracy) follows the order of articulatory complexity, and suggested that tone acquisition is closely related to the maturation of speech motor control.

In addition, the early acquisition of a particular feature such as affrication (e.g., [tɕ], [tɕʰ]) in Mandarin might highlight the possible influence of the ambient language on acquisition. Against nativist theory (Chomsky, 1965), the 'environmentalist' approach, originating from Skinner's (1957) behaviorism, considers children's learning as a stimulus–response process. This theory succeeds in drawing attention to the role of the environment (e.g., language input) in acquisition. Chen and Kent (2010) used a text corpus of 1,177, 984 Chinese characters (Cheng, 1982; Liu et al., 1975), including over 900,000 syllables with both consonant and vowel components. They found that one of the major characteristics of Mandarin is a slightly higher frequency of affricates (26.89%) in comparison to that of fricatives (24.97%). Before affricates are completely acquired, they can be replaced either by stops (e.g., [ts] by [t]) or by other affricates (e.g., [tsʰ] and [tɕʰ] by [ts]), but they have never been found to be replaced by fricatives. However, fricatives are sometimes replaced by affricates. For example, fricative [ɕ] was found to be substituted by the affricates [tɕ] and [tɕʰ] (Hua & Dodd, 2000; Hua, 2002). Moreover, the predominant production of [A] over other vowels in Mandarin-learning infants is closely related to the pattern in surrounded child-directed speech. The occurrence of low-central vowels (i.e., [A]) was over 28% (the highest frequency of all the Mandarin vowels) in caregivers' child-directed speech (Chen & Kent, 2010).

Hua and Dodd (2000) systematically studied the production of Mandarin among one- to four-year-old children and reported that children acquired phonological elements in the following order: tones were acquired first, followed by vowels and syllable-final consonants, which were then followed by syllable-initial consonants. The *phonological saliency hypothesis* (Hua & Dodd, 2000) might account for the order of phonological production in Mandarin, with Mandarin tones being the most salient factor. Mandarin tone is compulsory for every syllable. Switching lexical tones causes a change of word meaning, and there are only four choices. Syllable-initial consonants have the lowest saliency: their presence is optional (not all syllables have syllable-initial consonants), and there is a range of 21 syllable-initial phonemes that can be used. Vowels are compulsory syllable components. However, the relatively large number of options (including monophthongs, diphthongs, and triphthongs) lowers their saliency. In conclusion, differences in the saliency of individual components in a certain language may lead to variations in developmental patterns.

Singh and Fu (2016) proposed some other approaches to explain why Mandarin tones are acquired earlier than other speech units. One possible explanation for the early emergence of tones pertains to their provenance: vocal pitch. Pitch plays a role in every language in the form of intonation contrast, which is marked primarily by the pitch movements as well. For example, questions and exclamatory sentences are mainly distinguished by pitch across tonal and non-tonal languages alike. Similarly, the use of pitch-to-signal emotion is robust across various languages (Lieberman, 1967). Moreover, the centrality of pitch in auditory processing is evidenced at various stages in development. In the earliest phases of auditory perception, pitch, together with stress and rhythm, is preferentially available to infants before birth (Fifer & Moon, 1988).

In conclusion, children's phonological acquisition is a highly complicated process influenced by a variety of factors, including the role of different phonetic units within a given language and their relationship with other languages, the influence of surrounding speech environments, and the development of biological bases and cognitive ability in children. Consequently, no single theory is adequate to account for all the phenomena documented in studies of phonological acquisition; yet each can account for some aspects of the data (Stoel-Gammon & Sosa, 2007). The explicit dividing line between different theories is not always easy to determine because researchers borrow a feature of one theory and incorporate it into another, as the knowledge of phonological development evolves with time.

12.4 The Phonological Maturation of Mandarin Speech

The progressive development of phonological acquisition involves three stages: phonological emergence, phonological stabilization, and phonological maturation. Previous studies have mainly focused on the emergence and stabilization of different phonetic units of speech. The age of emergence records the first time that a child can articulate a phonetic unit and the age of stabilization indicates when a child can produce a phoneme with a certain degree of phonological accuracy and consistency (i.e., 66.7% accuracy). Unsurprisingly, the age of stabilization in most cases is later than the age of emergence. The question naturally arising is when children obtain the same speech competence as adults, in other words, the timing of phonological maturation. On the one hand, perceptual maturation refers to a perceptual competence of children which is equal to that of adult perceivers (e.g., Chen et al., 2017; Lee et al., 2012; Xi et al., 2009). On the other hand, production maturation can be evaluated either through the acoustic measurement (e.g., Chen, 2007; Ma, Chen, Wu, & Zhang, 2018; Shi & Wen, 2007) or through native adults' perceptual judgement of children's speech outputs (e.g., Wong, 2012, 2013).

On the acquisition of Mandarin tones, early studies (e.g., Chao, 1951; Clumeck, 1980; Hua, 2002; Hua & Dodd, 2000; Li & Thompson, 1977) indicated that lexical tones were produced early with considerable accuracy before the age of three. Judgments of tone errors in these studies were typically made by native adult observers in a categorical fashion (correct or incorrect). However, it is important to note that the aforementioned early production of tones in children before three years of age does not mean that these children have the same tonal production ability as adults. Innovations in speech analysis tools have afforded greater precision in the evaluation of early tone production. A series of studies (Wong, Schwartz, & Jenkins, 2005; Wong, 2012, 2013) have reported that three- to five-year-old preschoolers have not yet fully mastered the production of Mandarin tones. Tone productions of adults (control group) and children were collected in a picture-naming task and low-pass filtered to remove lexical information. Native speakers categorized the target tones in the low-pass filtered productions in which only tone information was reserved. Children's tone accuracy was compared to that of the adults to determine the level of mastery and developmental changes. None of the Mandarin tones produced by the three- to five-year-old children reached adult-like accuracy, suggesting a protracted course of development extending beyond age five. These findings stood in contrast to earlier studies that claimed very early acquisition (emergence or stabilization) of stable tone productions (Chao, 1951; Clumeck, 1980; Hua, 2002; Hua & Dodd, 2000; Li & Thompson, 1977).

On Mandarin tone perception, although a study of three-year-old children suggested that they already achieved relatively high perceptual accuracy of all four Mandarin tones (Wong et al., 2005), research on the developmental course of categorical perception (CP) of Mandarin tones is still limited. The study of CP is useful because it offers a much more refined perceptual method for tracing the course of the stabilization and maturation of children's fine-grained perception of Mandarin tones beyond age three. A higher degree of CP likely indicates enhanced perceptual ability in young children. Using the classic paradigm of CP Chen et al. (2017), explored how CP of Mandarin tones emerges among 70 four- to seven-year-old children and 16 adults (control group). Mandarin-speaking children exposed to a native tonal language could perceive Mandarin Tone 1 and Tone 2 categorically. The positions of the identification boundaries did not differ significantly between children and adults,

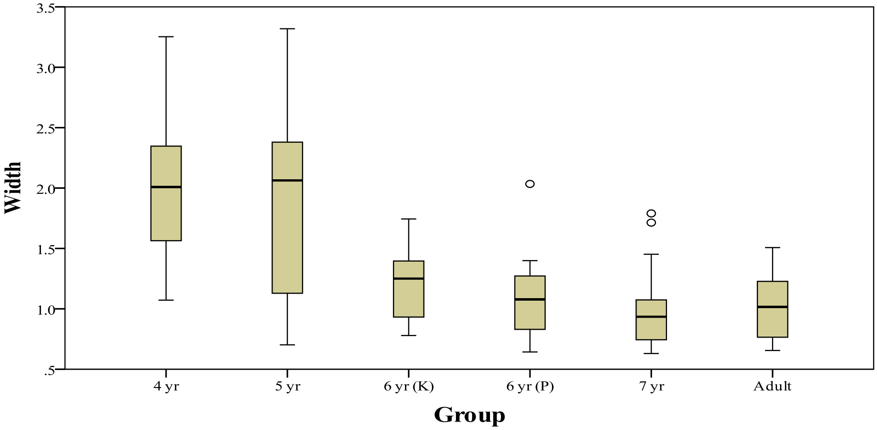

Fig. 12.2 Box plots of boundary widths within each age group, adapted from Chen et al. (2017)

but the boundary widths between Tone 1 and Tone 2 did differ significantly, with much narrower boundary widths (i.e., sharper boundaries) occurring in six-year olds than five-year olds (see Fig. 12.2). Moreover, with age, the ability to distinguish more fine-grained tonal differences of between-category pairs improved gradually due to perceptual accumulation. These findings contribute to the literature in discovering the general developmental course of CP during the maturation of tone perception in young children. The indexes of boundary width (i.e., transition slope) and between-category discrimination accuracy can thus be utilized quantitatively to explore the developmental trajectory of perceptual ability during the process of maturation.

Previous studies have mainly analyzed Mandarin consonant and vowel acquisition using conventional phonetics. In most cases, only one native adult, usually the author, did the transcription and then rated the speech naturally produced by children. The rating results were often subjective and to some extent susceptible to perceptual classification. Improvements in the acoustic analysis may help us objectively evaluate children's production and/or perception, and further track the maturational stage. For example, Shi and Wen (2007) compared the formant patterns of different Mandarin vowels spoken by young children and adults. The vowel spaces (defined by F1 and F2 values) of adults offer a standard F1–F2 graph to track the development of vowel production. Voice onset time (VOT) has also been utilized to show the development of production (Chen, 2007) and perception (Xi et al., 2009) of Mandarin aspirated vs. unaspirated stops. Chen (2007) showed that children go through a period (1;5–1;6) in which no distinction is made in the VOT of the stop consonants, passing to a stage (1;7–1;9) in which a systematic but not adult-like distinction is made, before reaching a final stage (1:10–2:11) in which the VOT values of stops resemble the adult model via a process of systematic refinement in VOT.

12.5 Future Directions in Mandarin Phonological Acquisition

Through discussion of previous studies on Mandarin phonological acquisition, this chapter examined the developmental trajectories (phonological emergence, stabilization, and maturation) of individual phonetic units of speech (e.g., Mandarin consonants, vowels, or tones). However, large discrepancies were also reported in the order and age of acquisition. Conflicting results can be attributed to many factors, including the criteria used, collection approach (cross-sectional or longitudinal research), speech mode (spontaneous production or imitation), sample size, age range of subjects, living area of subjects, number of transcribers, and so on. Unsurprisingly, these differences, especially in the criteria used, result in the identification of different ages and orders of acquisition of speech sounds. In order to establish a more systematic, scientific, reliable, and representative 'normative data set' describing the order and age of Mandarin phonological acquisition, future studies should pay close attention to controlling influential factors. Moreover, the normative data obtained in the future should be based on a large, representative sample in order to reflect the true population and therefore minimize individual differences.

Previous studies on Mandarin phonological acquisition focused mainly on the development of speech production. Studies investigating the perceptual development of Mandarin consonants and vowels are still few. Moreover, speech perception and production are presumed to be correlated constructs, exemplified by the fact that English speakers with a more fine-grained discrimination of phonetic contrast are also likely to produce the same phonemes with a greater degree of acoustic contrast (e.g., Fox, 1982). The extent of linguistic transfer between perception and production points to an important area of research in first language development (Singh & Fu, 2016). A more precise picture of the nature of the Mandarin production-perception interface will inform and deepen our understanding of how these two domains may be linked during Mandarin phonological acquisition.

The concern with generalizations about the order of acquisition leaves no room for considering the nature of individual differences in phonological development. Yet any careful comparison of different children learning the same language shows differences in the individual's paths of development. Some individual, cultural, and social factors, such as gender, socioeconomic status, sibling status, intelligence, personality, cognitive style, and parenting behaviors (in particular their language habits), have been studied in relation to phonological development (for a review, see Winitz, 1969). Discovering the variables that play an important role in children's language acquisition will have implications for how norms should be derived and applied to a clinical population (Dodd et al., 2003). Future studies may specifically focus on evaluating the influences of these individual and social factors on Mandarin phonological acquisition. It is also meaningful to investigate the specific influence of L2 language environment (another language or another Chinese dialect) on the age and order of Mandarin sound acquisition.

Moreover, behavioral research often requires overt responses and sustained concentration, which is relatively difficult for young children and infants. The development of neuroscience techniques offers us a valuable chance to uncover the neural substrates of how young brains process the smallest building blocks of speech (i.e., phonemes). Noninvasive techniques that enable the examination of language processing in infants and young children have advanced rapidly, including electroencephalography (EEG)/event-related potentials (ERPs), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and near-infrared spectroscopy (NIRS). Research suggested that exposure to language in the first year of life begins to set the neural architecture that supports infants' subsequent acquisition of language (Kuhl & Rivera-Gaxiola, 2008). Recently, a few ERP studies investigated the brain responses to different Mandarin phonetic units of speech (Cheng et al., 2013; Lee et al., 2012; Liu et al., 2014; Lee & Cheng, Chap. 7 this volume). These studies used mismatch responses (MMRs), consisting of mismatch negativity (MMN) and positive mismatch response (p-MMR), as possible brain indices to investigate the development of Mandarin speech perception. The developmental patterns revealed by various noninvasive techniques are important for understanding the underlying neural and physiological bases of Mandarin speech development.

Finally, as suggested by Wang (1978), how children learn a language—the transmission across generations—is clearly one of the vital questions in the whole of language change. As in historical sound change, both lexical and phonetic parameters are involved during the phonological acquisition in children. This point, often missed by most studies, helps connect microhistory, as seen in language acquisition, to mesohistory, as studied by historical linguistics. Two important studies in this area (Ferguson and Farwell, 1975; Hsieh, 1972) investigated the development of phonological production in relation to the acquisition of words, and showed that there was a primacy of lexical learning during phonological development. Importantly, the child does not progress by acquiring units like phonemes or allophones, but rather by gradually adding lexical items to his/her repertoire. The same sound appearing in different words may undergo different developmental trajectories, and the unity of the phoneme only emerges when the acquisition process is fully completed. Consequently, the basic unit of acquisition is something like the word, and many valuable conclusions would be derived from observing the lexically gradual nature of sound substitution during the long process of children's language learning.

References

Benoît, C., & Le Goff, B. (1998). Audio-visual speech synthesis from French text: Eight years of models, designs and evaluation at the ICP. *Speech Communication, 26*(1), 117–129.

Cao, J. X. (2003). One case study of early phonological development in Mandarin-speaking children. In *Proceedings of the 6th National Symposium on Modern Phonetics*, Tianjin, Tianjin normal University. [曹井香. 汉族儿童早期语音发展个案研究. 第六届全国现代语音学学术会议论文集. 2003].

Chao, Y. R. (1951). The Cantian idiolect: An analysis of the Chinese spoken by a twenty-eight-months-old child. In W. J. Fischel (Ed.), *Semitic and oriental studies* (pp. 27–44). Berkeley, CA: University of California Press.

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language, 44*(6), 1413–1434.

Chen, F. Y. (2007). *The development of voicing contrast in Mandarin: A longitudinal case study (Unpublished MA dissertation).* College of Foreign Languages, Hunan University. [陈霏燕. 普通话儿童塞音送气对立获得研究:一个个案研究. 湖南大学硕士学位论文, 2007].

Chen, L. M., & Kent, R. D. (2009). Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language, 36*(1), 73–84.

Chen, L. M., & Kent, R. D. (2010). Segmental production in Mandarin-learning infants. *Journal of Child Language, 37*(2), 341–371.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, C. M. (1982). Analysis of present-day Mandarin. *Journal of Chinese Linguistics, 10,* 281–358.

Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., & Lee, C. Y. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology, 38*(5), 281–300.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clumeck, H. V. (1977). *Studies in the acquisition of Mandarin phonology (Unpublished doctoral dissertation).* Berkerley: University of California.

Clumeck, H. (1980). The acquisition of tone. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Production* (Vol. 1, pp. 257–275). New York, NY: Academic Press.

Dodd, B. (1995). *Differential diagnosis and treatment of children with speech disorder*. London: Whurr Publishers.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617–643.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27,* 315–330.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language, 51,* 419–439.

Fifer, W. P., & Moon, C. (1988). Auditory experience in the fetus. In W. P. Smotherman & S. R. Robinson (Eds.), *Behavior of the fetus* (pp. 175–188). Caldwell, NJ: Telford Press.

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica, 39*(1), 1–22.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language, 55*(1), 37–58.

Hsieh, H. I. (1972). Lexical diffusion: Evidence from child language acquisition. *Glossa, 6*(1), 89–104.

Hsu, J. (1987). *A study of the various stages of development and acquisition of Mandarin Chinese by children in Taiwan milieu* (Unpublished MA dissertation). College of Foreign Languages: Fu Jen Catholic University.

Hua, Z. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children* (Vol. 3). Clevedon: Cromwell.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*(1), 3–42.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universal*. The Hague: Mouton.

Jeng, H.-H. (1979). The acquisition of Chinese phonology in relation to Jakobson's law of irreversible solidarity. In *Proceedings of the 9th International Congress of Phonetic Sciences* (Vol. 2, pp. 155–161). Copenhagen: University of Copenhagen.

Kent, R. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.
Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annual Review of Neuroscience, 31,* 511–534.
Lee, C. Y., Yen, H. L., Yeh, P. W., Lin, W. H., Cheng, Y. Y., Tzeng, Y. L., & Wu, H. C. (2012). Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia, 50*(14), 3228–3239.
Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(2), 185–199.
Lieberman, P. (1967). *Intonation, perception and language.* Cambridge, MA: MIT Press.
Lin, Y. H. (1989). *Auto-segmental treatment of segmental processes in Chinese phonology (Unpublished Ph. D. dissertation).* Austin: University of Texas.
Liu, C. Y. (2007). *The development of Mandarin phonology from 18 to 23 months in Shanghai (Unpublished MA dissertation).* College of Humanities and Communication, Shanghai Normal University. [刘春燕. 18–23个月儿童普通话的语音发展 (上海地区). 上海师范大学硕士学位论文, 2007].
Liu, H. M., Chen, Y., & Tsao, F. M. (2014). Developmental changes in mismatch responses to Mandarin consonants and lexical tones from early to middle childhood. *PLoS ONE, 9*(4), e95587.
Liu, I. M., Chuang, C. J., & Wang, S. C. (1975). *Frequency count of 40,000 Chinese words.* Taipei: Lucky Books Company.
Locke, J. (1983). *Phonological acquisition and change.* New York, NY: Academic Press.
Ma, J., Chen, X., Wu, Y., & Zhang, L. (2018). Effects of age and sex on voice onset time: Evidence from Mandarin voiceless stops. *Logopedics Phoniatrics Vocology, 43*(2), 56–62.
Norman, J. (1988). *Chinese.* Cambridge: Cambridge University Press.
Olmsted, D. (1971). *Out of the mouth of babes.* The Hague: Mouton.
Prather, E. M., Hedrick, D. L., & Kern, C. A. (1975). Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders, 40*(2), 179–191.
Shi, F., & Wen, B. Y. (2007). Vowel development in Mandarin-speaking children. *Zhongguo Yuwen, 2007*(5), 444–454. [石锋, 温宝莹. 汉语普通话儿童的元音发展. 中国语文, *2007*(5), 444–454].
Shiu, H.-S. (1990). *The phonological acquisition by Mandarin-speaking children: A longitudinal case study on children from 9 months through three years old (Unpublished MA thesis).* Taiwan Normal University.
Si, Y. Y. (2006). Mandarin phonological acquisition: A case study. *Contemporary Linguistics, 2006*(1), 1–16. [司玉英. 普通话儿童语音习得的个案研究. 当代语言学, *2006*(1), 1–16].
Singh, L., & Fu, C. S. (2016). A new view of language development: The acquisition of lexical tone. *Child Development, 87*(3), 834–854.
Skinner, B. F. (1957). *Verbal behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
Slobin, D. I. (1985). *The crosslinguistic study of language acquisition: Theoretical issues* (Vol. 2). Hillsdale, NJ: Psychology Press.
Stern, C., & Stern, W. (1907). *Monographien über die seelische Entwicklung des Kindes* (Vol. 1). Leipzig: J. A. Barth.
Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children.* Baltimore, MD: University Park Press.
Stoel-Gammon, C., & Sosa, A. V. (2007). *Phonological development*. Oxford: Blackwell.
Tsao, F. M. (2008). The effect of acoustical similarity on lexical tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology, 50*(2), 111–124.
Wang, W. S. Y. (1973). The Chinese language. *Scientific American, 228,* 50–60.
Wang, W.S.-Y. (1978). The three scales of diachrony. In B. B. Kachru (Ed.), *Linguistics in the seventies* (pp. 63–75). Urbana, IL: Department of Linguistics, University of Illinois.
Winitz, H. (1969). *Articulatory acquisition and behavior*. New York, NY: Appleton-Century-Crofts.
Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141–151.

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America, 133*(1), 434–443.

Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research, 48*(5), 1065–1079.

Wu T. M., Xu, Z. Y. (1979). A preliminary analysis of language development in children during the first three years. *Acta Psychologia Sinica, 11*(2), 153–165. [吴天敏, 许政援. 初生到三岁儿童言语发展记录的初步分析. 心理学报, *11*(2), 153–165].

Xi, J., Jiang, W., Zhang, L. J., & Shu, H. (2009). Categorical perception of VOT and lexical tones in Chinese and the developmental course. *Acta Psychologia Sinica, 41,* 572–579.

Xie, H. (2009). *The acquisition of Mandarin initial consonants (Unpublished MA dissertation).* College of Foreign Languages, Hunan University. [谢衡. 汉语普通话儿童的声母习得研究. 湖南大学硕士学位论文, 2009].

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61–83.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123–139.

Zhong, X. (2013). *A case study of consonant acquisition in Mandarin-speaking children (Unpublished MA dissertation).* College of Foreign Languages, Tianjin Normal University. [钟心. 普通话儿童辅音习得个案研究. 天津师范大学硕士学位论文, 2013].