Chapter 2
The Phonetic Realizations of the Mandarin Phoneme Inventory: The Canonical and the Variants

Janice Fon

Abstract This chapter provides an overview on the phonetic realizations of the Mandarin phoneme inventory. There are two major sections. The first is a general description of the phoneme inventory, which includes five vowels and 19 consonants, along with four lexical tones. In addition to acoustic properties of individual phonemes, major allophonic rules are also discussed. The second section covers some variations on consonants, vowels, and tones in three major Mandarin varieties of Taiwan, Singapore, and China. Some variations are fairly region-specific, while others are more commonly found across various dialects. The former includes the retroflexed vowel suffix in the Mainland variety, the qualitative difference in realizing the neutral tone between the Taiwan and the Mainland variety, and the so-called fifth tone in Singapore Mandarin. The latter includes the deretroflexion and hypercorrection of sibilants, both of which can be found in Taiwan and Singapore Mandarin, and the syllable-final nasal mergers, which can be found in all three major dialects. Interestingly, these cross-dialectal variations also show large within-region variabilities. Both the canonical and the variant realizations of the phonological system are discussed in light of child language acquisition.

Mandarin is by far the most widely spoken language in the world. As of 2013, it boasts an estimated L1 population close to 900 million (Lewis, Simons, & Fennig, 2016), accounting for more than 12% of the total world population at the time (cf. Population Reference Bureau, 2013). It is the sole official language of Taiwan and China, and is one of the four official languages in Singapore. This chapter contains two main sections. The first is devoted to an overview of the phonetic realizations of Mandarin phonology, and the second focuses on some variations in the three major Mandarin variants of Taiwan, Singapore, and China. All three main aspects of phonology, i.e., vowels, consonants, and tones, are discussed.

J. Fon (✉)
Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan
e-mail: jfon@ntu.edu.tw

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_2

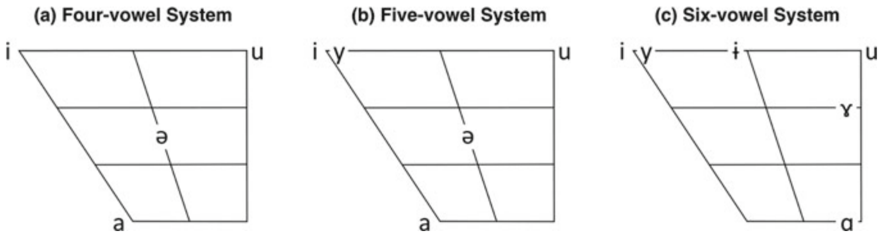

Fig. 2.1 Three different proposals for the Mandarin vowel system

2.1 The Phoneme Inventory

2.1.1 Vowels

Depending on the theoretical frameworks, Mandarin is said to have a system ranging from four to six single vowels. The four-vowel system is based on feature geometry and was proposed by Wu (1994) (Fig. 2.1a), while the six-vowel system is based on phonemic evidence and was proposed by C.-C. Cheng (1973) (Fig. 2.1c). Most other scholars argued for a five-vowel system, with an inventory of three high vowels, one mid-vowel, and one low vowel (Chao, 1968; R. L. Cheng, 1966; Duanmu, 2007; Y.-H. Lin, 1989; Wan & Jaeger, 2003), as shown in Fig. 2.1b. As all three proposals include largely overlapping sets of vowels, this chapter follows the perspective of the majority and reviews only the five-vowel system in detail.

The three high vowels in Mandarin could be nicely demonstrated by a minimal triplet, as shown in (1).[1] These three phonemes behave in a rather similar fashion and become a homorganic glide when they precede a non-high vowel, as shown in (2). [i] and [u] can also be the second part of a diphthong and act as an off-glide, as shown in (3).

(1)	*í* /i/	*yú* /y/	*wú* /u/
	‘aunt’	‘fish’	‘nil’

(2)	*yǎn* /ian/→[jan]	*wǎn* /uan/→[wan]	*yuǎn* /yan/→[ɥan]
	‘to perform’	‘bowl’	‘faraway’

(3)	*nǎi* /nai/→[naɪ]	*nǎo* /nau/→[naʊ]
	‘milk’	‘brain’

[1] *Hanyu Pinyin* is adopted for Mandarin romanization throughout this chapter.

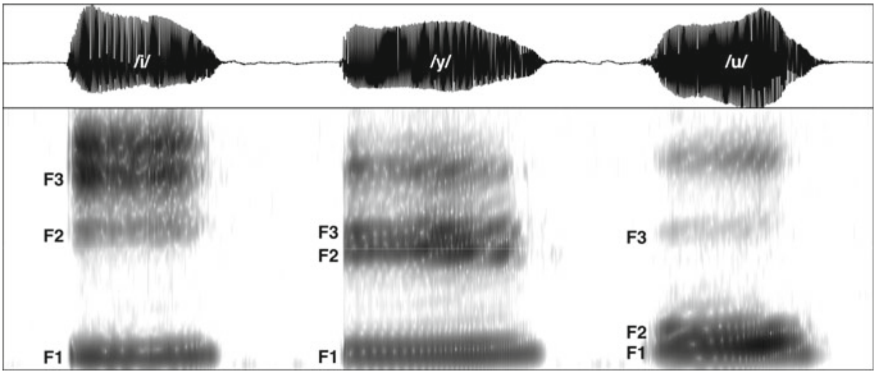

Fig. 2.2 Waveforms and spectrograms of the three high vowels, *yí* /i/ 'aunt', *yú* /y/ 'fish', and *wú* /u/ 'nil'. The first three formants are labeled accordingly

Figure 2.2 shows the waveforms and spectrograms of the three high vowels, *yí* /i/ 'aunt', *yú* /y/ 'fish', and *wú* /u/ 'nil', produced by a female speaker. Notice that the acoustic cues correspond nicely with the articulatory descriptions of the vowels. All three vowels are high, and therefore their F1s are unanimously low. /i/ and /y/ are both front vowels, while /u/ is a back vowel, and therefore the former two have higher F2s than the latter. /y/ and /u/ are both rounded vowels and therefore have relatively low F3s (and F2s), as compared to the unrounded /i/.

The high vowel /i/ is worth further mention. Besides being realized as a high front vowel and a glide, it also has a third allophone [ɨ] that occurs after dental and retroflex sibilants [ts ts^h s $tʂ^h$ $ʂ^h$ ʐ] (see below) (C.-C. Cheng, 1973). (4) contains a near-minimal triplet of [i] and [ɨ],[2] and their waveforms and the spectrograms are shown in Fig. 2.3. Compared to their counterpart [i] in *tì* [ti] 'ground', the F2s in *zì* [tsɨ] 'Chinese character' and *zhì* [tʂɨ] 'mole (face)' are much lower, indicating backing of the tongue position.

(4)	*tì* /ti/→[ti]	*zì* /tsi/→[tsɨ]	*zhì* /tʂi/→[tʂɨ]
	'ground'	'Chinese character'	'mole (face)'

[2]The phone has been problematic in Mandarin phonology and has invited much discussion and debate. Classical views usually argued for two phones instead of one. For example, Norman (1988) proposed two apical vowels of [ɿ] and [ʅ], Lee-Kim (2014) argued for two syllabic approximates of [ɹ̩] and [ɻ̩], and Duanmu (2007) suggested two syllabic fricatives of [z̩] and [ʐ̩]. The former ones of each pair are designated to occur after dental sibilants, while the latter ones are designated to occur after retroflexes. However, since the formants in Fig. 2.3 show little frication, implying that it is at least not always realized as a syllabic fricative, and since the major difference between the two renditions of the phone (*zì* [tsɨ] 'Chinese character' vs. *zhì* [tʂɨ] 'mole (face)') mainly lies in F3, not F2, implying that there is little change in tongue backness, this chapter adopts a unifying symbol [ɨ] instead, following C.-C. Cheng (1973), and views the phonetic variation as a product of coarticulation. There is a slight departure from C.-C. Cheng's (1973) original proposal, however, as this chapter argues for a nonphonemic status instead (see Fig. 2.1).

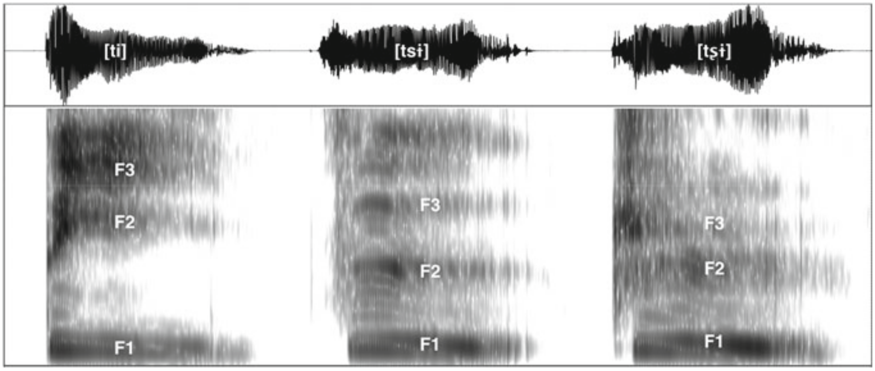

Fig. 2.3 Waveforms and spectrograms comparing the two allophonic variants of /i/ in *tì* [ti] 'ground', *zì* [tsɨ] 'Chinese character', and *zhì* [tʂɨ] 'mole (face)'. The first three formants are labeled in white

Mandarin only has one mid-vowel, but its frontness and rounding show environmentally dependent allophonic changes while its height stays the same. Most previous works agree that there are at least four variants, [ə], [e], [o], and [ɤ] (C.-C. Cheng, 1973; R. L. Cheng, 1966; Y.-H. Lin, 1989). [ə] occurs in CVN syllables and is considered the default (R. L. Cheng, 1966; Y.-H. Lin, 1989; Wan & Jaeger, 2003; Wu, 1994). [e] occurs in CjV and CɥV syllables, while [o] occurs in CwV syllables. Finally, [ɤ] occurs in CV syllables. (5) shows a set of examples.

(5)	*lèng* /ləŋ/→[ləŋ]	*liè* /liə/→[lje]	*luò* /luə/→[lwo]	*lè* /lə/→[lɤ]
	'absent-minded'	'to crack'	'to fall'	'happy'

Figure 2.4 shows the waveforms and spectrograms of the four allophonic variants of the mid-vowel /ə/ in *lèng* [ləŋ] 'absent-minded', *liè* [lje] 'to crack', *luò* [lwo]'

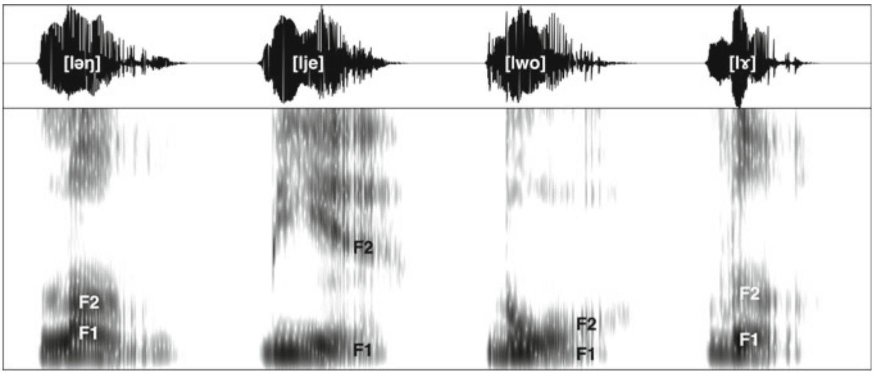

Fig. 2.4 Waveforms and spectrograms of the four allophonic variants of the mid-vowel /ə/ in *lèng* [ləŋ] 'absent-minded', *liè* [lje] 'to crack', *luò* [lwo] 'to fall', and *lè* [lɤ] 'happy'. The first two formants of the main vowels are labeled in black or white accordingly

to fall', and *lè* [lɤ] 'happy'. Notice that both [e] and [ɤ] are mid-high vowels and therefore have similar F1s. [e] has a higher F2 than [ɤ] because it is a front vowel. [o] and [ɤ] have similar F1s and F2s because they are both mid-high back vowels, but the former has a slightly lower F2 due to rounding. Compared with [e] and [o], [ə] has a higher F1, showing that it is relatively low in height, and its F2 value is intermediate between that of [e] and [o], showing that it is in the central position.

Similarly, Mandarin also only has one low vowel, and its frontness and height are context-dependent and show allophonic changes. However, its unrounded quality stays the same. Most studies agree that there are at least three allophonic variants, [a], [ɑ], and [ɛ] (C.-C. Cheng, 1973; R. L. Cheng, 1966; Y.-H. Lin, 1989; Wu, 1994). [a] is default and occurs in open and C(w)Vn syllables (R. L. Cheng, 1966; Y.-H. Lin, 1989; Wan & Jaeger, 2003). [ɑ] occurs in CVu and C(G)Vŋ, and [ɛ] occurs in CjVn and CɥVn syllables. (6) is a set of examples.

(6)	*án* /an/→[an]	*áng* /aŋ/→[ɑŋ]	*ián* /ian/→[jɛn]
	'1.SG (regional)'[4]	'to raise high'	'salt'

[4] 1.sg: first person singular pronoun

Figure 2.5 shows the waveforms and spectrograms of the three allophonic variants of the low vowel /a/ in *án* [an] '1.SG', *áng* [ɑŋ] 'to raise high', and *ián* [jɛn] 'salt'. Notice that both [a] and [ɛ] are front vowels and therefore have higher F2s than the back vowel [ɑ]. [ɛ] also has a lower F1 than [a] and [ɑ], indicating its higher vowel height.

There is an additional retroflex vowel /ɚ/ in Mandarin, which can only occur by itself in a bare V syllable (Duanmu, 2007). (7) shows a minimal pair of /ɚ/ versus /ə/. Figure 2.6 shows a spectrographic comparison between the two vowels. Notice that because of the rhotic quality of the vowel /ɚ/, the third formant is drastically lowered as compared to that of the mid-vowel /ə/.

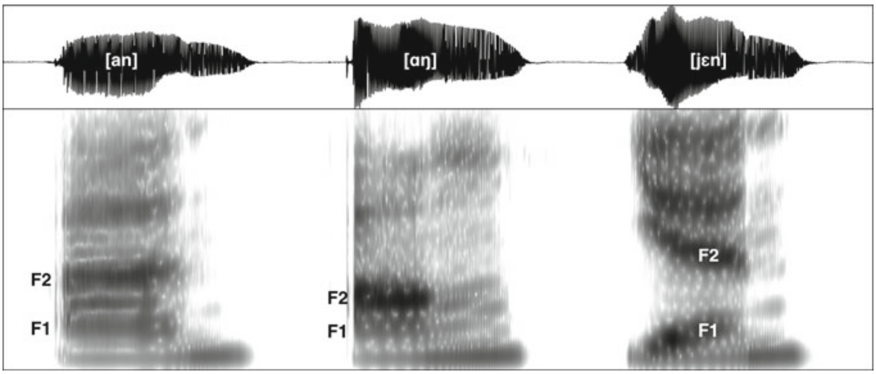

Fig. 2.5 Waveforms and spectrograms showing the three allophonic variants of the low vowel /a/, *án* [an] '1.SG', *áng* [ɑŋ] 'to raise high', and *ián* [jɛn] 'salt'. The first two formants of the main vowels are labeled in black or white accordingly

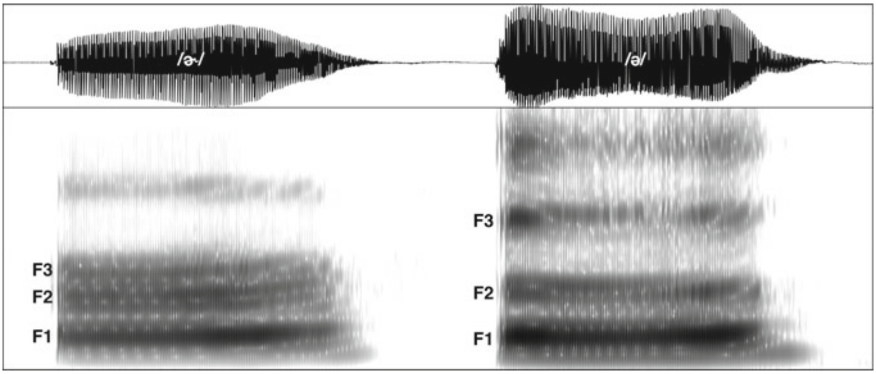

Fig. 2.6 Waveforms and spectrograms of the retroflex vowel *ér* /ɚ/ ‘son’ and the mid-vowel é /ə/ ‘goose’. The first three formants of the main vowels are labeled accordingly

Table 2.1 Mandarin consonant chart

	Labial		Dental		Retroflex		Velar	
Stop	p	p^h	t	t^h			k	k^h
Affricate			ts	ts^h	tʂ	tʂ^h		
Fricative	f		s		ʂ	ʐ	x	
Nasal	m		n				ŋ	
Lateral			l					

(7) *ér* /ɚ/ *é* /ə/

‘son’ ‘goose’

2.1.2 Consonants

Mandarin has 19 consonants, including 6 stops, 4 affricates, 5 fricatives, 3 nasals, and 1 lateral (Chao, 1968; R. L. Cheng, 1966; Duanmu, 2007) (Table 2.1).[3] The stops occupy three places of articulation, labial, dental, and velar, and have both the aspirated and the unaspirated series. (8) shows a minimal sextuple. The waveforms and the spectrograms of the sextuple are shown in Fig. 2.7.

---

[3]Some researchers contended that Mandarin consonant inventory also includes an additional set of three alveolo-palatal sibilants /tɕ tɕ^h ɕ/ [e.g., Luo (1993)]. However, since these three are in complementary distribution with the velar series /k k^h x/, the dental series /ts ts^h s/, and the retroflexes /tʂ tʂ^h ʂ/, this chapter views them as allophones of other phonemes and discusses them in a later section.

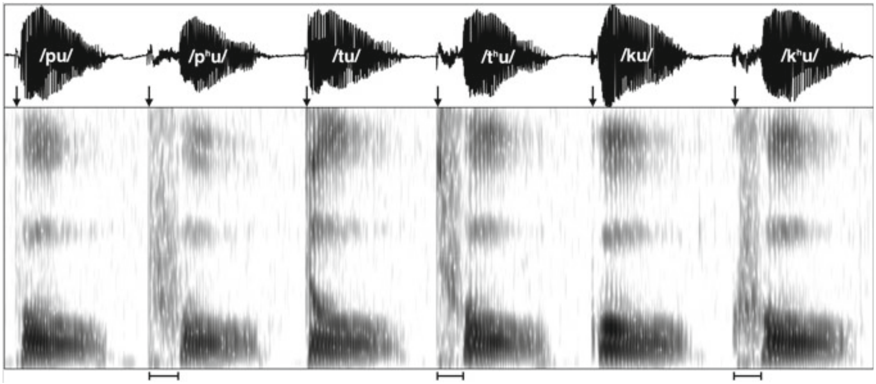

Fig. 2.7 Waveforms and spectrograms of the six stops, *bŭ* /pu/ 'to mend', *pŭ* /pʰu/ 'sheet music', *dŭ* /tu/ 'to gamble', *tŭ* /tʰu/ 'soil', *gŭ* /ku/ 'ancient', and *kŭ* /kʰu/ 'bitter'. The downward arrows above the spectrograms indicate the stop bursts, and the horizontal line segments beneath the spectrograms indicate the aspiration noise of aspirated stops

(8)	*bŭ* /pu/	*pŭ* /pʰu/
	'to mend'	'sheet music'
	dŭ /tu/	*tŭ* /tʰu/
	'to gamble'	'soil'
	gŭ /ku/	*kŭ* /kʰu/
	'ancient'	'bitter'

There are two sets of affricates in Mandarin, the dental and the retroflex. Like their stop counterparts, both the aspirated and the unaspirated series are included. (9) shows a minimal quadruple. The waveforms and the spectrograms of the quadruple are shown in Fig. 2.8. Notice that the lowest major energy concentration for the frication portion of the affricates is lower for the retroflexes than that for the dentals.

(9)	*zàn* /tsan/	*càn* /tsʰan/
	'to praise	'bright'
	zhàn /tʂan/	*chàn* /tʂʰan/
	'to stand'	'to shiver'

Fricatives are the largest set of inventory in Mandarin. In total, there are three sibilants and two nonsibilants. Of the three sibilants, there is one voiceless dental /s/, one voiceless retroflex /ʂ/, and one voiced retroflex /ʐ/. (10) shows a minimal triplet. The waveforms and the spectrograms of the triplet are shown in Fig. 2.9. Like the affricates, the retroflexes have lower major energy concentration than the dental.

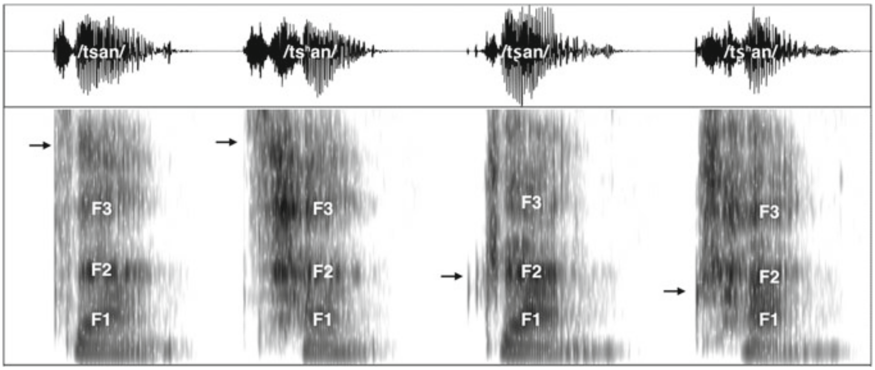

Fig. 2.8 Waveforms and spectrograms of the four affricates, *zàn* /tsan/ ‘to praise’, *càn* /ts^{h}an/ ‘bright’, *zhàn* /tʂan/ ‘to stand’, and *chàn* /tʂ^{h}an/ ‘to shiver’. The rightward arrows indicate the lowest major energy concentration for the noise burst of the affricates. The first three formants of the main vowels are labeled in white accordingly

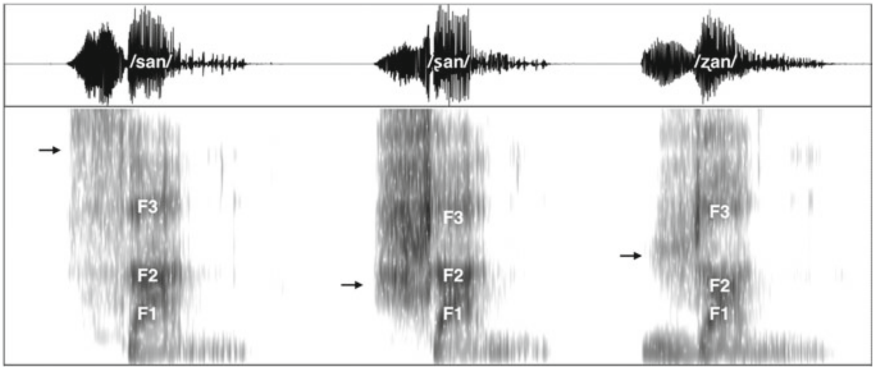

Fig. 2.9 Waveforms and spectrograms of the three Mandarin sibilant fricatives, *sǎn* /san/ ‘umbrella’, *ǎn* /ʂan/ ‘to flash’, and *rǎn* /ʐan/ ‘to dye’. The rightward arrows indicate the lowest major energy concentration for the fricative noise. The first three formants of the main vowels are labeled in white accordingly

(10)	*sǎn* /san/	*shǎn* /ʂan/	*rǎn* /ʐan/
	‘umbrella’	‘to flash’	‘to dye’

The two nonsibilant fricatives are labial /f/ and velar /x/. (11) shows a minimal pair. The waveforms and the spectrograms of the pair are shown in Fig. 2.10.

(11)	*fú* /fu/	*hú* /xu/
	‘blessing’	‘lake’

Mandarin has three nasals, /m n ŋ/, and one liquid /l/. /m n l/ can occur in the onset position, as indicated in (12). The waveforms and the spectrograms of the triplet are shown in Fig. 2.11. /n ŋ/ can occur in the coda position, as shown in (13).

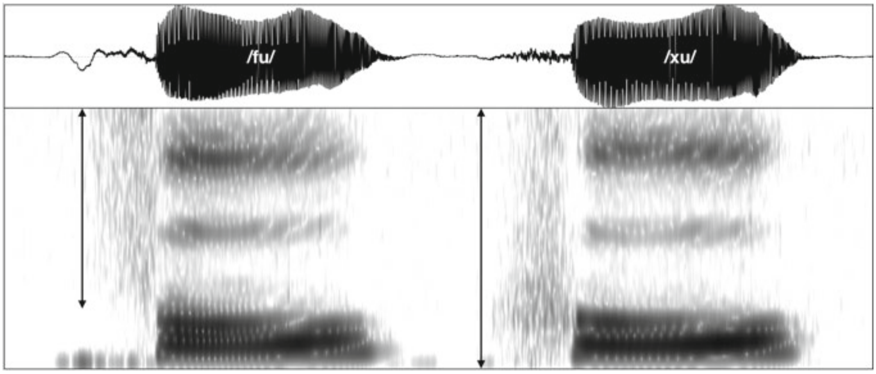

Fig. 2.10 Waveforms and spectrograms of the two nonsibilant fricatives, *fú* /fu/ 'blessing' and *hú* /xu/ 'lake'. The up–down arrows indicate the range of frequency noise for the two fricatives

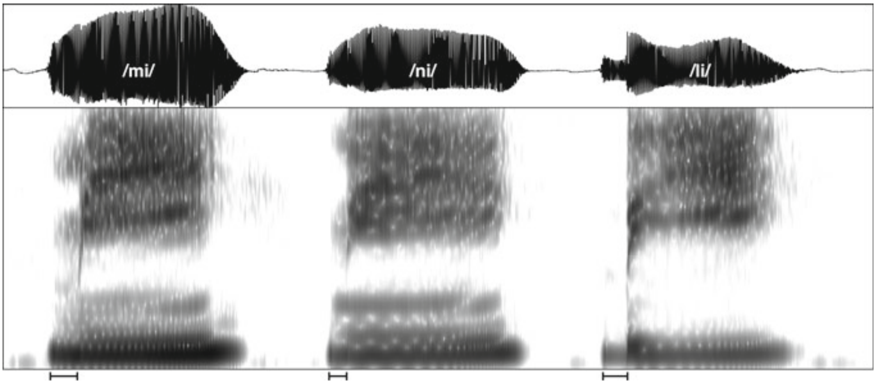

Fig. 2.11 Waveforms and spectrograms of the three onset sonorants, *mí* /mi/ 'mystery', *ní* /ni/ 'mud', and *lí* /li/ 'pear'. The horizontal line segments beneath the spectrograms indicate the sonorant sections

As illustrated in Fig. 2.12, the two nasals differ not only in nasal formants, but also in the F2 and F3 offsets of the preceding vowels. /n/ has both a falling F2 and a falling F3, while /ŋ/ has a rising F2 and a falling F3.

(12)	*mí* /mi/	*ní* /ni/	*lí* /li/
	'mystery'	'mud'	'pear'
(13)	*pín* /pʰin/	*píng* /pʰiŋ/	
	'poverty'	'level'	

There is also a special set of alveolo-palatal sibilants [tɕ tɕʰ ɕ], which are in complementary distribution with the velars /k kʰ x/, the dentals /ts tsʰ s/, and the retroflexes /tʂ tʂʰ ʂ/. The alveolo-palatal set only occurs before /i y/, while the other

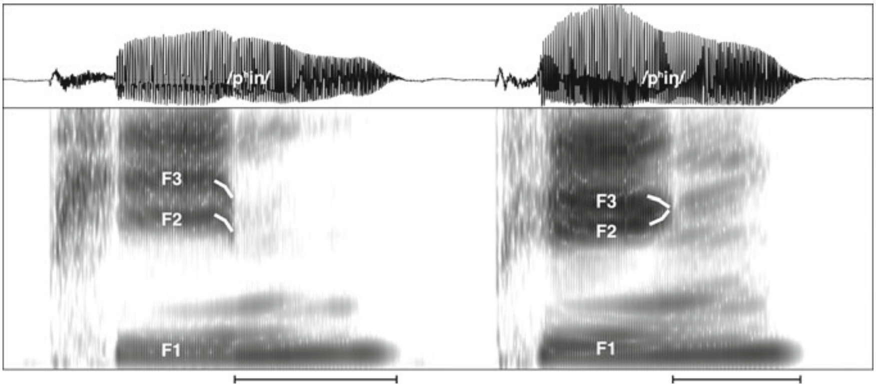

Fig. 2.12 Waveforms and spectrograms of the two nasal codas, *pín* /pʰin/ 'poverty' and *píng* /pʰiŋ/ 'level'. The first three formants of the main vowels are labeled in white. The F2 and F3 offset tracings of the main vowels are also shown in white. The horizontal line segments beneath the spectrograms indicate the nasal sections

three only occur elsewhere, as shown in (14). Diachronically, the alveolo-palatal set stems from two historical sources, the velars and the dentals. Synchronically, some scholars have identified it with the velar set [e.g., Chao (1968), R. L. Cheng (1966)], while others have identified it with the dental set [e.g., Hartman (1944)]. In reality, the evidence could go either way (Duanmu, 2007; Y.-H. Lin, 1989). Figure 2.13 shows the waveforms and the spectrograms of an alveolo-palatal triplet.

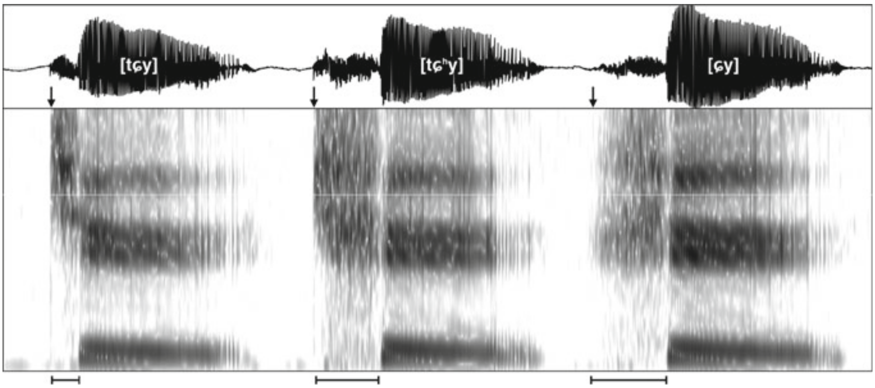

Fig. 2.13 Waveforms and spectrograms of the three alveolo-palatals, *jŭ* [tɕy] 'to lift', *qŭ* [tɕʰy] 'to take', and *xŭ* [ɕy] 'to promise'. The downward arrows above the spectrograms indicate the stop bursts, and the horizontal line segments beneath the spectrograms indicate frication noise

(14)	*jǐ* [tɕi]	*qǐ* [tɕʰi]	*xǐ* [ɕi]
	'to squeeze'	'to rise'	'to wash'
	jǔ [tɕy]	*qǔ* [tɕʰy]	*xǔ* [ɕy]
	'to lift'	'to take'	'to promise'
	*/ki-/ */ky-/ */tsi-/ */tsy-/ */tʂi-/ */tʂy-/	*/kʰi-/ */kʰy-/ */tsʰi-/ */tsʰy-/ */tʂʰi-/ */tʂʰy-/	*/xi-/ */xy-/ */si-/ */sy-/ */ʂi-/ */ʂy-/

2.1.3 Tones

There are four citation tones in Mandarin, traditionally characterized as Tone 1, Tone 2, Tone 3, and Tone 4. Tone 1 is a high-level tone, Tone 2 is a rising tone, Tone 3 is a dipping (i.e., falling–rising) tone, and Tone 4 is a falling tone (Chao, 1968). (15) shows an example of a minimal quartet, using a fairly common transcription convention of suffixing the tonal number to the IPA transcription of the syllable (Duanmu, 2007). Figure 2.14 shows an example of the F0 tracks of the four tones. Although tones are realized by continuous pitch contours, different sections of the contours seem to have different weightings for different tones. Lee & Wiener (Chap. 3 of this volume) and Tsai & Liu (Chap. 10 of this volume) give a good overview of

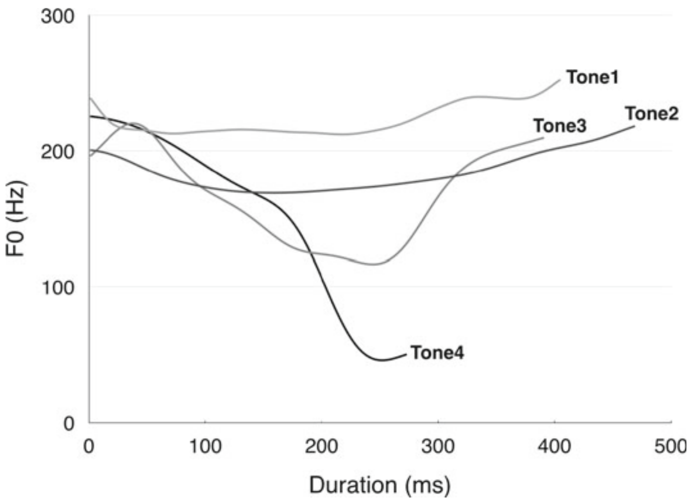

Fig. 2.14 F0 tracks of the four tones in Mandarin on the syllable *ying* /iŋ/

how these contours are perceived by novice (infants and nontone language users) and expert listeners (Mandarin adults).

(15) *yīng* /iŋ1/ 'eagle' *yíng* /iŋ2/ 'to win' *yǐng* /iŋ3/ 'shadow' *yìng* /iŋ4/ 'hard'

When a word contains two consecutive Tone 3s, the first Tone 3 undergoes a change and transforms into a Tone 2, making it homophonous with a Tone 2 + Tone 3 sequence. This is known as the Tone 3 sandhi rule. For example, *mǎ-liǎn* /ma3-lian3/ 'oblong face (lit. "horse face")' becomes homophonous with *má-liǎn* /ma2-lian3/ 'pockmarked face' because the former undergoes the Tone 3 sandhi rule (16). Although the rule is classically posed as obligatory (Chao, 1968), the actual realization is far more complex and variable, as it has been shown that the rule interacts intricately with both linguistic factors like word frequency (Yuan & Chen, 2014) and nonlinguistic factors like speech rate (H.-B. Lin, 1982).

(16) *mǎ-liǎn* /ma3-lian3/→[ma2-lian3]
'oblong face (lit. 'horse-face')'

má-liǎn /ma2-lian3/→[ma2-lian3]
'pockmarked face'

Not all syllables bear the four tones mentioned above. Only stressed syllables do. For unstressed syllables, they have the neutral tone (i.e., Tone 0) instead. (17) shows an example of a minimal pair. By destressing the second syllable, the meaning changes from 'wife and son' to 'wife' only. Traditionally, neutral tones are considered to be a tonal category that does not have an intrinsic tonal value of its own, and its tonal realization is largely dependent on the preceding tone (Chao, 1968). However, Chen and Xu (2006) claimed that the neural tone owns a mid-level tonal target, but is qualitatively different from non-neutral tones in that the articulatory strength for reaching the target is relatively weak and inefficient, and thus the speed at which such a target is approached is relatively slow. As a consequence, neutral tones are more susceptible to coarticulatory forces of surrounding tones. In either case, neutral tones are realized to be of shorter duration and weaker amplitude, as shown in Fig. 2.15.

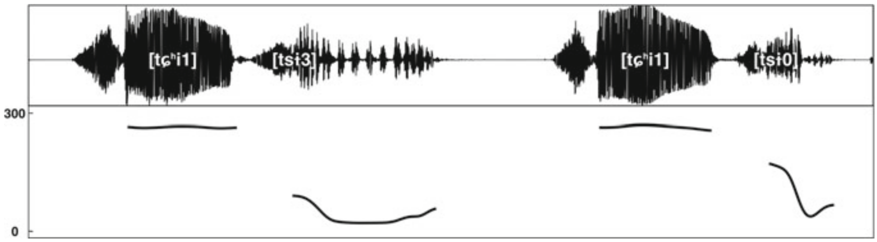

Fig. 2.15 Waveforms and F0 contours for the minimal pair of *qī-zǐ* [tɕʰi1-tsɨ3] 'wife and son' and *qī-zi* [tɕʰi1-tsɨ0] 'wife'

(17) *qī-zǐ* [tɕʰi1-tsɨ3] *qī-zi* [tɕʰi1-tsɨ0]

'wife and son' 'wife'

2.1.4 Syllable Structure

Mandarin syllables have a maximal form of CGVX plus tone (Duanmu, 2007). Except for the nucleus vowel slot, all the others are optional. VX constitutes the rime, and the X slot can either be fulfilled by an off-glide of a diphthong or a nasal. The status of G is more controversial. Traditionally, it is analyzed as part of the final, which includes a medial slot reserved for G, and the rime (R. L. Cheng, 1966) (Fig. 2.16a). However, more recently, Duanmu (2007) claimed that G should be grouped together with C to form the onset, as it adds a secondary articulation to C when both are present (Fig. 2.16b). Figure 2.17 shows a spectrographic comparison between Mandarin *suì* [sʷeɪ] 'to shatter' and English *sway* [sweɪ]. Notice that the major energy concentration for [s] in *suì* is much lower than that for [s] in *sway* as a result of more lip protrusion in the former. In addition, there is a clearer (and longer) voice portion of [w] in *sway* than that in *suì*. These crosslinguistic differences imply that Mandarin glides are qualitatively different from English glides when they are positioned in consonant clusters and might be more adequately characterized as a secondary articulation to the leading consonant rather than as an independent phone instead.

As there are 18 consonants possible for the onset C slot (i.e., all consonants except for /ŋ/, see Table 2.1), 3 glides possible for the G slot, 5 vowels possible for the V slot (excluding the retroflex vowel /ɚ/, see Fig. 2.1b), and 4 sounds (/i, u, n, ŋ/) possible for the coda X slot, and everything except for the V slot is optional, there could theoretically be 19 (C slot) × 4 (G slot) × 5 (V slot) × 5 (X slot) = 1900 syllables

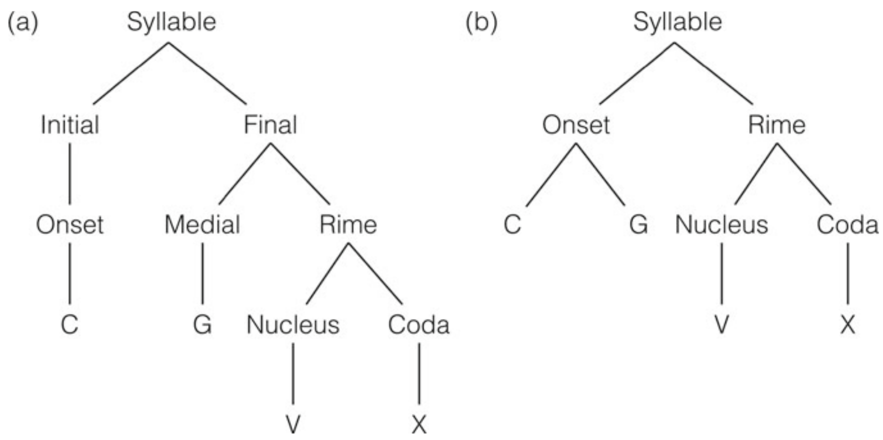

Fig. 2.16 Two different views of the Mandarin syllable structure. Please see text for explanation

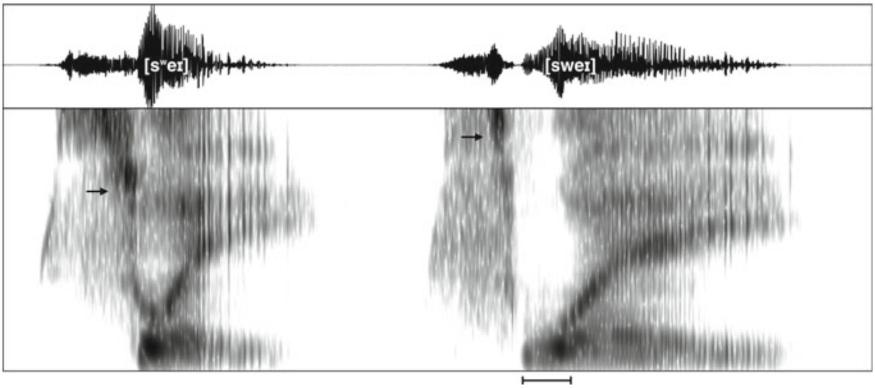

Fig. 2.17 Waveforms and spectrograms for Mandarin *suì* [s^{w}eɪ] 'to shatter' (left) and English *sway* [sweɪ]. The rightward arrows indicate the lowest major energy concentration for the fricative noise of [s]. The horizontal line segments beneath the spectrogram indicate the section for the glide [w]. Please see text for explanation

possible in Mandarin, disregarding tones. However, in reality, there are only about a few more than 400 syllables existing in the inventory (Duanmu, 2007). Even when tones are included, the number only goes up to about 1300, which is far fewer than the number of monosyllables in English. Although this inevitably results in a large number of homophonous monosyllabic morphemes, homophony is not a prevalent issue in daily communication, as monosyllabic words only constitute about 27% of the lexicon, and the majority (≈ 70%) is in fact bisyllabic (He & Li, 1987). If contextual information is further taken into consideration, miscommunication due to homophony is virtually nonexistent.

In summary, Mandarin is much like many of the East Asian languages, with a moderate segment inventory size and a relatively simple syllable makeup, both of which seem to superficially imply a smooth and easy acquisition process. However, the language is phonologically marked in several aspects. First, it is a tone language that contains both register and contour tones, along with complicated tone sandhi realization rules (Duanmu, 2007; H.-B. Lin, 1982; Yuan & Chen, 2014). As most of the more well-studied Indo-European languages are non-tonal (Maddieson, 2013), Mandarin could provide us with a better understanding of the role of tone and how it interacts with other syllabic components in the process of acquisition. Secondly, being a tone language, Mandarin shows special constraints on its implementation of prosody, as the acoustic correlates of prosody and tone overlap to a large extent (Peng et al., 2005). How the interaction of the two affects the acquisition process compared to other non-tonal languages is an interesting issue both in theory and in application. Finally, although there is nothing phenomenal regarding the consonant inventory size of Mandarin, it contains a special set of retroflex sibilants (i.e., /tʂ tʂ^{h} ʂ ʐ/) that is relatively rare among languages of the world. According to the UCLA Phonological Segment Inventory Database (Maddieson, 1984), out of the 451 languages investigated, only 30 of them incorporate at least one retroflex sibilant in their inventory, accounting for about 7%. If one only considers languages that have at least as many as four retroflex sibilants like Mandarin, the number goes down drastically to only 6, which is a mere 1%. Thus, studying Mandarin could also provide new insights on how crosslinguistically common and uncommon sounds are acquired in the first few beginning years.

2.2 Variations in Phoneme Realization

As Mandarin is widely spoken by a vast population of speakers, variation in its phonological implementation is understandably inevitable. Since variation likely adds further complications to the already formidable task for young beginning speakers, description of such a phonological system would be incomplete if variation in phonological implementation is lacking. As Singapore has gained its independence from Malaysia in 1965, and Taiwan has been politically separated from China for nearly 70 years, it is not surprising that many of the noted variations in Mandarin phonology are found among the three standard Mandarin varieties. However, more recent studies also focus on some within-variety variations in Taiwan due to differential degrees of language contact with Southern Min, a local substrate with which about 70% of the Taiwan population show at least some familiarity (Huang, 1993). In this section, all three types of variations of vowels, consonants, and tones are delineated.

2.2.1 Vowels

One aspect of vowels that shows drastic dialectal variations is the retroflexed vowels in the Mainland variety. The retroflex vowel /ɚ/ can occur as a suffix signaling diminutiveness or achieving certain stylistic purposes. It assumes the coda position and oftentimes completely replaces the original coda of the host syllable, if any (Duanmu, 2007). (18) shows an example of a minimal pair. While the non-retroflexed *zhè* can mean both ‘this’ and ‘here’, the retroflexed *zhèr* is used only to indicate ‘here’. When a syllable is retroflexed, the vowel assumes a rhotic quality and its F3 is substantially lowered, as shown in Fig. 2.18. Although retroflexed vowels are quite common in the Mainland variety, it is virtually nonexistent in Singapore (Chew, 2002) and Taiwan. In our Taiwan Mandarin spontaneous speech corpus (Fon, 2004), which currently includes more than 50 h of transcribed monologues, there is not a single instance of retroflexed vowels found.[4] In other words, one can probably safely say this is a vocalic trait that is exclusive to the Mainland variety.

[4]The speakers in the Taiwan Mandarin corpus included both young (ages 20–35) and old speakers (ages 50–65) who spent all or most of their childhood and teenage years in the designated sampling locations.

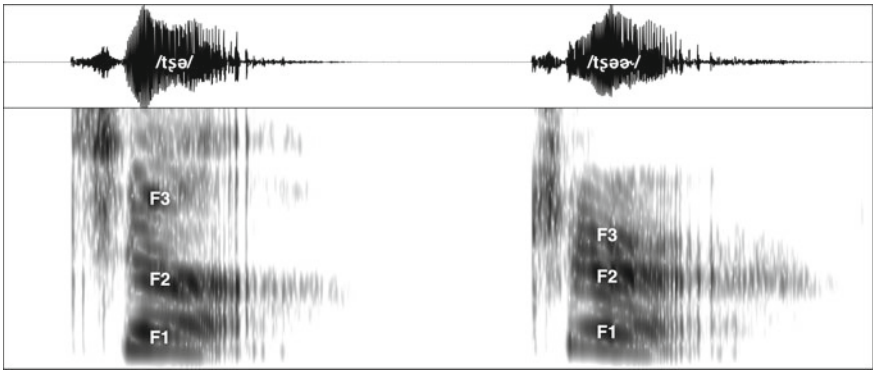

Fig. 2.18 Waveforms and spectrograms for *zhè* /tʂə/ 'this; here' and *zhèr* 'here' /tʂəɚ/. The first three formants are labeled accordingly

(18) *zhè* /tʂə/ *zhèr* /tʂəɚ/
'this; here' 'here'

2.2.2 Consonants

Dialectal variations are even more prevalent in consonants. In this section, two types of consonant variations are discussed. One is the realization of dental and retroflex sibilants, and the other concerns the syllable-final nasal mergers.

2.2.2.1 Dental and Retroflex Sibilants

Sibilant realization is extremely variable in Taiwan (Chuang, 2009; Chuang, Chiang, & Fon, 2012; Chuang, Wang, & Fon, 2015) and Singapore Mandarin (Ng, 1985) due to frequent contact with local language substrates. Two main processes have been documented. One is to realize the retroflex sibilants as dentals, appropriately termed the deretroflexion rule, and the other is the reverse of the deretroflexion rule, realizing the dental sibilants as retroflexes. Since the latter is deemed as a way speakers employ to counteract the prevalence of deretroflexion, it is usually termed the hypercorrection rule and is more commonly observed in formal styles (Chuang, 2009; Chung, 2006; Ng, 1985). As shown in (19), the two sets of rules between voiceless dentals and voiceless retroflexes are fairly straightforward, as they are the exact opposite of each other. However, realization for the voiced retroflex /ʐ/ is more variable. As it lacks a phonological dental counterpart (cf. Table 2.1), previous studies showed that its most common deretroflexed realization is [l] and [z] is only the second most common (Chan, 1984; Chuang et al., 2015).

(19) Deretroflexion: /tʂ tʂʰ ʂ/ → [ts tsʰ s]
/ʐ/ → [z l]

Hypercorrection: /ts tsʰ s/ → [tʂ tʂʰ ʂ]

Gender and genre are effective factors in determining the application of the rules (Chuang et al., 2012; Ng, 1985). Deretroflexion is more likely found among males than females, in spontaneous than read speech, and in connected speech than word lists. The trend is observed in both Taiwan and Singapore Mandarin (Fig. 2.19). On the other hand, more dialectal variation was found for hypercorrection. Although Singapore Mandarin still shows both a gender and a genre effect, as male speakers are more likely to apply hypercorrection than females, and spontaneous speech being the least likely to facilitate the application of this rule, similar trends are not observed in Taiwan Mandarin, as its hypercorrection rates are unanimously low (Fig. 2.19). In fact, both rules are far more prevalent in Singapore Mandarin than Taiwan Mandarin. Deretroflexion in Singapore Mandarin is nearly complete in male spontaneous speech, and its hypercorrection is almost 60% in male tongue twisters. In contrast, deretroflexion in Taiwan Mandarin is at best less than 40% for male spontaneous speech, and all hypercorrection rates regardless of gender and genre hover around 5% only. This might have been due to differential attitudes towards the rules in the two variants. Ng (1985) claimed that Singapore Mandarin speakers generally deem retroflex realization as undesirable, as they often associate it with foreigner speech and snobbishness. Most Singaporean speakers find it more natural with a variety that reveals local identity. On the other hand, most speakers from Taiwan associate deretroflexion as a stigma for being nonstandard (Chan, 1984; Chuang et al., 2012; Kubler, 1985a, 1985b) and would thus avoid doing so, at least consciously, when situation arises. Since hypercorrection is meant to counteract deretroflexion, it is possible that Singaporeans find it more necessary to do so in a formal context than Taiwan speakers as the former shows a much higher deretroflexion rate than the latter.

At least for the Mandarin variety in Taiwan, sibilant realization is also affected by prominence (Chuang, 2009; Chuang & Fon, 2010). As shown in Fig. 2.20, sibilants in more prominent positions are realized with a larger degree of dentalization. This is true for sibilants of both dental and retroflex place, and also for both original and derived sibilants (i.e., those that underwent deretroflexion/hypercorrection). This implies that retroflex sibilants at a prosodically prominent position can be acoustically similar to dental sibilants at a prosodically weaker position, and perceptual ambiguity might potentially arise when one is devoid of context. The figure also shows that there is a gender difference in the maintenance of the two sibilant places. Female speakers in general maintain a more distinct phoneme space for the dental and retroflex series, respectively. Similar levels of distinction hold for both original and derived sibilants. On the other hand, male speakers not only show less distinction between original dentals and retroflexes, and between derived dentals and retroflexes, and their derived sibilants are also not as close to the intended targets as their female counterparts. This indicates that although male Taiwan Mandarin speakers may not have as much overt mixing between retroflex and dental sibilants at the phonemic level as their

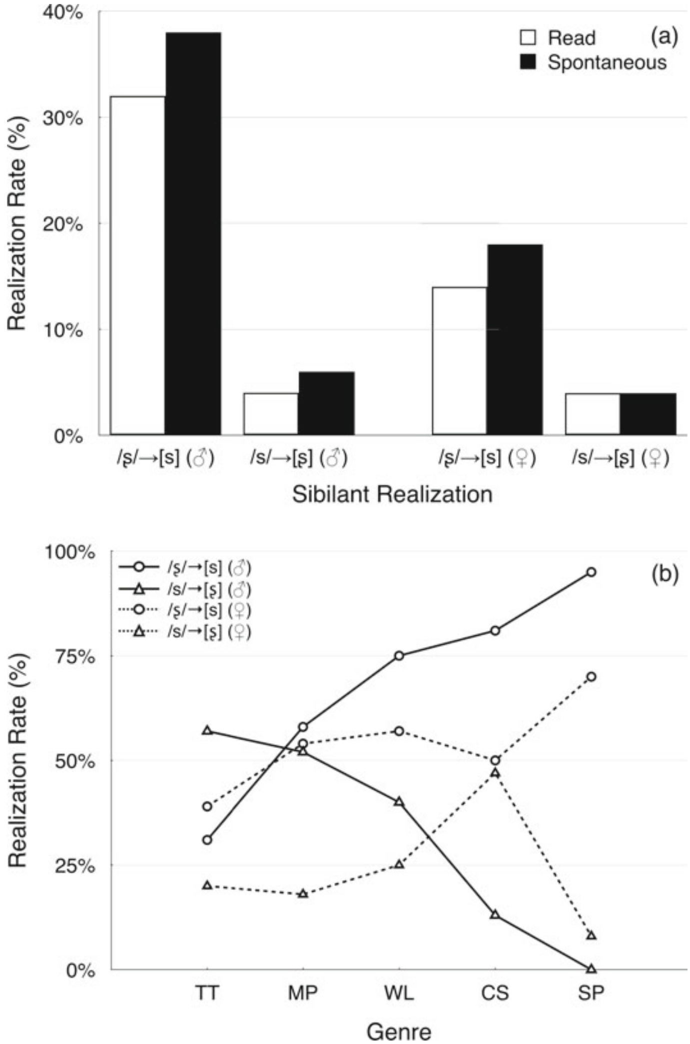

Fig. 2.19 Realization rates of deretroflexion /ʂ/ → [s] and hypercorrection /s/ → [ʂ] in **a** Taiwan (Chuang et al., 2012) and **b** Singapore Mandarin (Ng, 1985). TT: tongue twisters; MP: minimal pairs; WL: word lists; CS: connected speech; and SP: spontaneous speech

Singaporean counterparts (cf. Figure 2.19), their phonetic realization of the sibilants indeed shows that the two series are becoming more like each other.

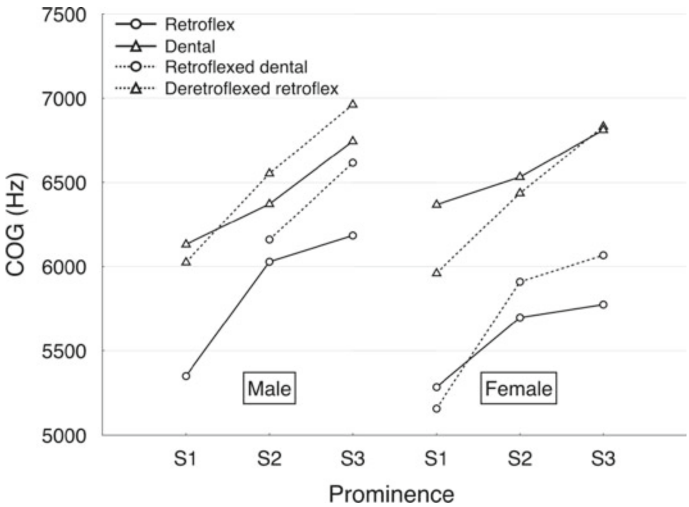

Fig. 2.20 Effect of prominence on the realization of voiceless sibilants. Being negatively correlated with the degree of retroflexion, the center of gravity (COG) was taken from the middle 10 ms of the fricative spectrum for all six voiceless sibilants /tʂ tʂʰ ʂ ts tsʰ s/ using a spontaneous speech corpus (Chuang, 2009; Chuang & Fon, 2010). Prominence was defined using the stress tier in the pan-Mandarin-ToBI system (Peng et al., 2005). S2 is the default level, S3 shows extra prominence, and S1 indicates unstressed positions. Retroflex: retroflex sibilants that are underlyingly retroflex, /tʂ tʂʰ ʂ/ → [tʂ tʂʰ ʂ]; dental: dental sibilants that are underlyingly dental, /ts tsʰ s/ → [ts tsʰ s]; retroflexed dental: dental sibilants that underwent hypercorrection, /ts tsʰ s/ → [tʂ tʂʰ ʂ]; deretroflexed retroflex: retroflex sibilants that underwent deretroflexion, /tʂ tʂʰ ʂ/ → [ts tsʰ s]. There were no unstressed male tokens with the retroflexed dental in the data collected

2.2.2.2 Syllable-Final Nasal Mergers

The two Mandarin coda nasals /n/ and /ŋ/ are allowed to occur after all five vowels. However, they are found to be merging with each other after /i/ and /ə/ (C.-Y. Chen, 1991; Fon, Hung, Huang, & Hsu, 2011; Yang, 2010). As shown in (20), merging for the former is bidirectional, and /in/ → [iŋ] and /iŋ/ → [in] are both observed, while merging for the latter is unidirectional, and only /əŋ/ → [ən] is found.

(20) After /i/: *liú-xīn* /liou-ɕin/→[ljoʊ-ɕiŋ]
‘to be careful’

liú-xīng /liou-ɕiŋ/→[ljoʊ-ɕin]
‘shooting star’

After /ə/: *xué-shēng* /ɕyə-ʂəŋ/→[ɕɥe-ʂən]
‘student’

The three nasal mergers are not equally favored in the three major Mandarin varieties. As indicated in Fig. 2.21a, Singapore Mandarin predominantly uses /in/ → [iŋ] and /əŋ/ → [ən], and few /iŋ/ → [in] were found (C.-Y. Chen, 1991). On the other hand, Mainland Mandarin mainly adopts /in/ → [iŋ] and /iŋ/ → [in], while /əŋ/ → [ən] was rarely observed (Yang, 2010). Studies did not agree on how nasal mergers work in Taiwan Mandarin. C.-Y. Chen (1991) claimed that it uses /in/ → [iŋ] and /əŋ/ → [ən], but no /iŋ/ → [in], while Yang (2010) argued that it uses /əŋ/ → [ən] and /iŋ/ → [in], but no /in/ → [iŋ] instead. Although this controversy seems

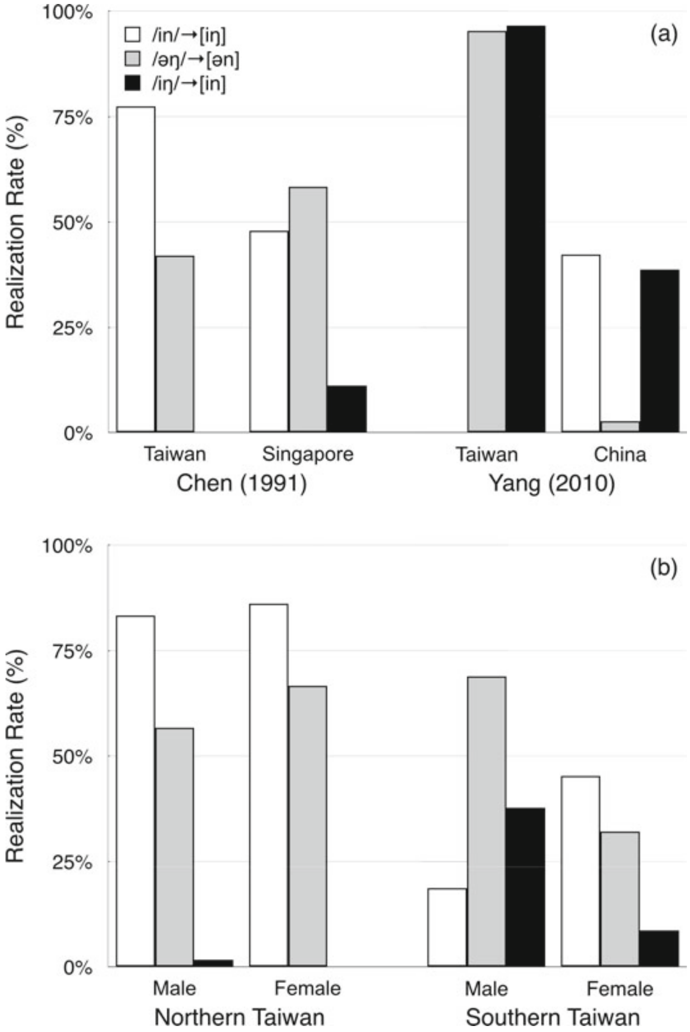

Fig. 2.21 Nasal mergers found in **a** Taiwan, Singapore, and China (C.-Y. Chen, 1991; Yang, 2010), and in **b** northern and southern Taiwan (Fon et al., 2011)

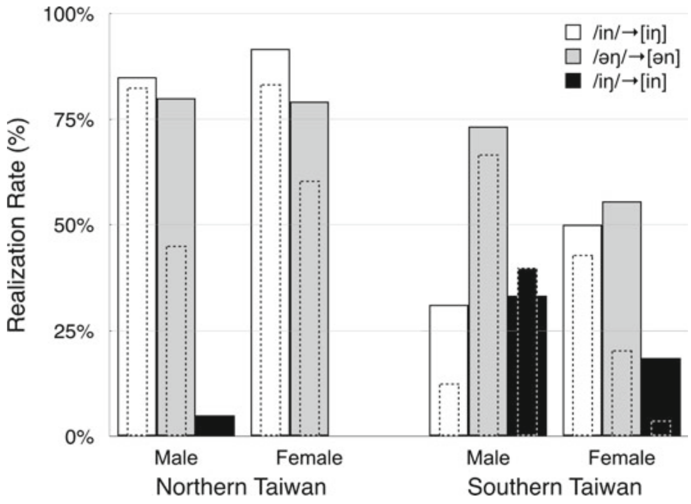

Fig. 2.22 Effect of context on the application of nasal mergers in Taiwan Mandarin. Solid bars represent target syllables in sentence-final positions, and dashed bars represent syllables in isolation (Fon et al., 2011)

intriguing, Fon et al. (2011) showed that this discrepancy is likely due to dialectal variations within the variety. As indicated in Fig. 2.21b, northern Taiwan speakers predominantly use /in/ → [iŋ] and /əŋ/ → [ən], much like what was found in C.-Y. Chen (1991), while southern Taiwan speakers could potentially use all three, which at least partially corroborated with Yang's (2010) findings.

Like the realization of voiceless sibilants, the application of nasal mergers can be context-dependent, at least in Taiwan Mandarin (Fon et al., 2011). As shown in Fig. 2.22, although context exerts little effect for /in/ → [iŋ] among northerners and southern females, the effect is more robust for southern males, and the latter group tends to apply the merger more in a sentential context. On the other hand, /əŋ/ → [ən] showed an opposite trend, and the context effect is more robust for northerners and southern females than southern males. As Taiwan Mandarin speakers generally regarded /in/ → [iŋ] as a form that is more prestigious than /əŋ/ → [ən] and /iŋ/ → [in], this implies that the application of nasal mergers is likely not only context-dependent, but also prestige- and therefore speaker-group-dependent, at least for the Mandarin variety in Taiwan.

2.2.3 *Tones*

Dialectal variations are also found in the realization of tones. This section focuses on two aspects of tonal variations, the realization of neutral tones, and that of the 'fifth tone' in Singapore Mandarin.

2.2.3.1 Realization of Neutral Tones

As mentioned above, Chao (1968) claimed that the actual pitch height of a neutral tone is largely dependent on the tonal value of the preceding syllable. It is realized the highest in pitch when it is preceded by a Tone 3, and the lowest when it is preceded by a Tone 4. Neutral tones following Tone 1 and Tone 2 are realized with mid-pitch. According to Chao (1968), neutral tones are more susceptible to coarticulatory forces due to lack of intrinsic tonal values and thus readily adopt the tonal values of their neighboring tones. However, Chao's (1968) account might not be viable in some varieties of Mandarin, as neutral tone realization seems to show dialect-dependent variations (Fig. 2.23). Although much variance is indeed found across different tonal contexts in the Mainland variety, neutral tones in Taiwan Mandarin are realized rather unanimously as mid-falling, with little effect from the preceding syllable being observed. This implies that the neutral tone in Taiwan Mandarin is qualitatively different from its Mainland counterpart, and may own an unswerving intrinsic tonal value of a mid-fall instead. Even if Chen and Xu's (2006) account for neutral tones is adopted, a defining difference for the neutral tones between the two dialects still exists. As mentioned earlier, Chen and Xu (2006) claimed that the fundamental difference between a neutral and a regular tone does not lie in the existence of an intrinsic tonal value, but rather, in the articulatory strength with which one exerts in reaching the tonal target. This may be an accurate description of neutral tones in Mainland Mandarin, but the consistent mid-fall realization of neutral tones in the Taiwan variety shows that there is no fundamental difference in the realization strength between a neutral and a regular tone in the dialect. In other words, the neutral tone in Taiwan Mandarin might have been more rightly termed as a 'fifth' tone instead, as it shares all of the characteristics of a regular tone except for being intrinsically short. On the other hand, the neutral tone in the Mainland variety still maintains a qualitative distinction from its regular tones and thus still fulfills the current categorization of being 'neutral'.

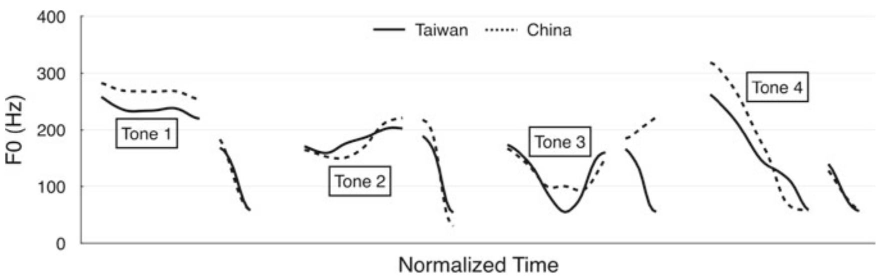

Fig. 2.23 Realization of neutral tones after a minimal quartet *ting-zhe* in Taiwan and Mainland Mandarin. Tone 1: *tīng-zhe* /tiŋ1-tʂə0/ 'listen ASP (listening)'; Tone 2: *tíng-zhe* /tiŋ2-tʂə0/ 'rest ASP (resting)'; Tone 3: *tǐng-zhe* /tiŋ3-tʂə0/ 'support ASP (supporting)'; Tone 4: *tìng-zhe* /tiŋ4-tʂə0/ 'to allow (one to act arbitrarily) ASP (allowing)' (ASP: aspect marker.)

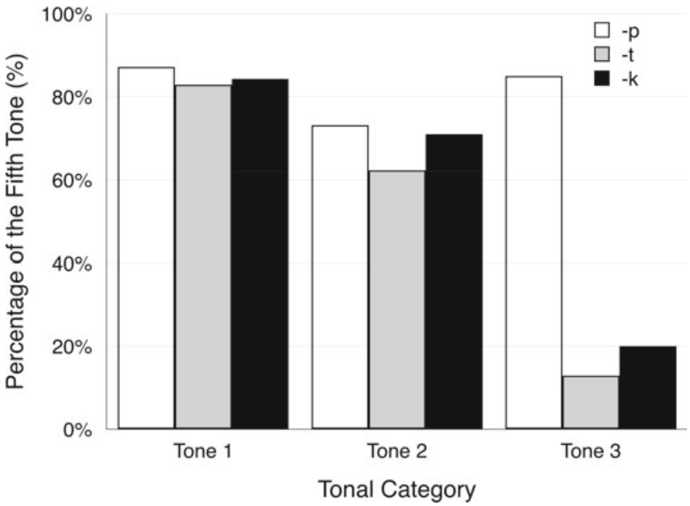

Fig. 2.24 Realization of the 'fifth tone' in Singapore Mandarin with regard to its tonal category in Modern Mandarin and its original entering tone endings (i.e., *-p*, *-t*, *-k*) in Middle Chinese. Data from C.-Y. Chen (1983)

2.2.3.2 'Fifth Tone' in Singapore Mandarin

Singapore speakers also tend to have an additional 'fifth tone' in their Mandarin, but the source of this fifth tone is completely different from the one that is suggested above for Taiwan Mandarin. According to C.-Y. Chen (1983), the fifth tone in Singapore Mandarin is characterized by a falling contour, resembling Mandarin Tone 4, but is often accompanied by a final glottal stop. When the final stop is present, the tone is short and intense; otherwise, the tone is barely distinguishable from a regular Mandarin Tone 4. The fifth tone only occurs in words that have etymologically an entering tone in Middle Chinese.[5] Since the majority of Singaporean Chinese speak at least one southern Chinese language natively and acquire Mandarin only after they enter the school system, C.-Y. Chen (1983) argues that the fifth tone might have stemmed from a negative transfer from the southern Chinese languages in Singapore, which still preserve entering tones in their phonology. Figure 2.24 shows that the actual realization of the fifth tone is very much dependent on the tonal category and etymology. Tone 1 shows the highest realization rate of over 80%, followed by Tone 2. For Tone 3 syllables, only those that had a *-p* ending in Middle Chinese show a high realization rate of the fifth tone. For syllables that previously ended in *-t* or *-k* in Middle Chinese, the realization rate is fairly low.

The variations mentioned above are by no means exhaustive. Rather, they are meant to merely serve as an illustration of the wide variety of Mandarin that Mandarin-speaking children might encounter during their acquisition process. Some

[5] An entering tone is a tone that occurs on syllables ending with /p t k ʔ/.

of the variations discussed are regionally distinct, as in the cases of the retroflex vowel suffix in the Mainland variety, the different ways of realizing neutral tones between Taiwan and Mainland Mandarin, and the fifth tone in Singapore Mandarin. From an acquisition point of view, this might not be as problematic, as long as Mandarin-learning children from different regions are considered and studied separately, and the data procured are analyzed within respective regional contexts. However, for variations that occur within a variety, such as the realization of voiceless sibilants in Taiwan and Singapore Mandarin, and the syllable final nasal mergers in Taiwan Mandarin, the situation might be much more complex. In order to master the system, children not only have to acquire the variable phonological rules, but also the appropriate occasions for applying them. Take the deretroflexion rule for example, for children acquiring Mandarin in Taiwan and Singapore, in addition to becoming proficient in producing the /ʈʂ $ʈʂ^h$ ʂ ʐ/ set, they also need to be familiar with the differential effect of various genres in order to determine whether the deretroflexed set should be used instead. For children acquiring Taiwan Mandarin, they need to carry this one step further and take note of the prosodic structure in which the voiceless sibilants occur so as to accurately determine the amount of deretroflexion to be applied in a native fashion. This is no doubt a fairly daunting task, and when placed in the context of first language acquisition, it raises several important issues. Although previous studies have shown that retroflexes are acquired relatively early in Mainland Mandarin (Li & Munson, 2016; Zhu & Dodd, 2000),[6] the complication of involving deretroflexion in the realization of the retroflex set in Taiwan and Singapore might substantially deter its acquisition process for Mandarin-learning children in these two locations. For these young learners of the language, they not only need to master the articulation of the retroflex set, as their counterparts in China do, but are also required to become proficient in applying the deretroflexion rule in contexts deemed to be adequate by adult natives in respective sites in order to be considered successful in acquiring the full usage of the retroflex set. How retroflexes are acquired in Taiwan and Singapore in specific and how variations across different varieties affect the acquisition process in general are relevant and significant to the Mandarin acquisition process and are thus worth further investigation.

Acquisition is a dynamic process that reflects not only the mentality of the learner (i.e., the child), but also the unique linguistic and social combination of the learned (i.e., the language). This is especially true in the case of Mandarin, as its wide geographical span and diverse speaker background inevitably entail large variability of both sides, which interact intricately with each other, fermenting into distinctive regional tangs. This chapter thus provides not only a broad introduction to various aspects of Mandarin phonology, but also includes both the commonalities and differences across the three major varieties of Mandarin (i.e., Taiwan, Singapore, and China). Even though this might have created a picture that is fuzzier than ideal, it is

[6]Zhu and Dodd (2000) reported that 75% of the children between ages 2;1 and 2;6 could produce/ʈʂ $ʈʂ^h$ ʂ/accurately at least once, and 75% of the children reach this criterion between ages 3;1 and 3;6 for/ʐ/. Li & Munson (2016) reported about 75% accuracy for/ʂ/ at around 3;6.

closer to reality, and thus one hopes that it could set a more realistic perspective for understanding the results in the later chapters.

References

Chan, H.-C. (1984). *The phonetic development of Mandarin / / in Taiwan: A sociolinguistic study.* (MA), Fu Jen Catholic University, Taipei.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chen, C.-Y. (1983). A fifth tone in the Mandarin spoken in Singapore. *Journal of Chinese Linguistics, 11*(1), 92–119.

Chen, C.-Y. (1991). The nasal endings and retroflexed initials in Peking Mandarin: Instability and the trend of changes. *Journal of Chinese Linguistics, 19*(2), 139–171.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech-evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica, 63*, 47–75.

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

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics, 2*(2), 135–158.

Chew, C. H. (2002). *Xinjiapo Huayu Cihui yu Yufa (Lexicon and syntax in Singapore Mandarin).* Singapore Lingzi Media.

Chuang, Y.-Y. (2009). *An acoustic study on voiceless retroflex and dental sibilants in Taiwan Mandarin spontaneous speech.* (M.A. thesis), National Taiwan University, Taipei.

Chuang, Y.-Y., Chiang, Y.-J., & Fon, J. (2012). *The effects of context and Min dialect on the realizations of / / variants in Taiwan Mandarin.* Paper presented at the 2nd Workshop on Sound Change, Kloster Seeon, Germany.

Chuang, Y.-Y., & Fon, J. (2010). *The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech.* Paper presented at the 5th International Conference on Speech Prosody.

Chuang, Y.-Y., Wang, S.-F., & Fon, J. (2015). *Cross-linguistic interaction between two voiced fricatives in Mandarin-Min simultaneous bilinguals.* Paper presented at the International Congress of Phonetic Sciences, Glasgow, Scotland, U.K.

Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication, 16*(2), 197–214.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). Oxford: Oxford University Press.

Fon, J. (2004). *A preliminary construction of Taiwan Southern Min spontaneous speech corpus* (NSC-92–2411-H-003–050-).

Fon, J., Hung, J.-M., Huang, Y.-H., & Hsu, H.-J. (2011). Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin. *Language and Linguistics, 12*(2), 273–311.

Hartman, L. M. (1944). The segmental phonemes of the Peiping dialect. *Language, 20*(1), 28–42.

He, K., & Li, D. (1987). *Xiandai Hanyu San Qian Changyong Ci Biao [Three Thousand Most Commonly Used Words in Modern Chinese].* Beijing: Beijing Shifan Daxue Chubanshe.

Huang, S. (1993). *Yuyan, Shehui yu Zuqun Yishi—Taiwan yuyan shehuixue de yanjiu [Language, society, and ethnic identity: Studies in language sociology in Taiwan].* Taipei: Crane Publishing.

Kubler, C. C. (1985a). *The development of Mandarin in Taiwan: A case study of language contact*. Taipei: Student Book.

Kubler, C. C. (1985b). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics, 27*(2), 156–176.

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

Li, F., & Munson, B. (2016). The development of voiceless sibilant fricatives in Putonghua-speaking children. *Journal of Speech and Hearing Research, 59*(4), 699–712.

Lin, H.-B. (1982). *Comparison of the differences in tone sandhi among slow speech, normal speech and fast speech in Mandarin Chinese.* (M.A. thesis), The Ohio State University, Columbus, OH, USA.

Lin, Y.-H. (1989). *Autosegmental treatment of segmental process in Chinese phonology.* (Ph.D. dissertation), University of Texas at Austin, Austin.

Luo, C.-C. (1993). *Kuoyuxue [Studies of Mandarin Chinese].* Taipei: Wu-Nan.

Maddieson, I. (1984). The UCLA phonological segment inventory database. Retrieved 4 June 2015. https://web.phonetik.uni-frankfurt.de/upsid.html.

Maddieson, I. (2013). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ng, B. C. (1985). A study of the variable /sh/ on Singapore Mandarin. In D. Bradley (Ed.), *Language policy, language planning and sociolinguistics in South-East Asia* (pp. 31–37). Canberra, Australia: Pacific Linguistics.

Norman, J. (1988). *Chinese.* Cambridge: Cambridge University Press.

Peng, S.-H., Chan, M. K. M., Tseng, C.-Y., Huang, T., Lee, O. J., & Beckman, E. M. (2005). Towards a pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230–270). New York: Oxford University Press.

Population Reference Bureau 2013. 2013 World Population Data Sheet. Population Reference Bureau.

Wan, I.-P., & Jaeger, J. (2003). The phonological representation of Taiwan Mandarin vowels: A psycholinguistic study. *Journal of East Asian Linguistics, 12*(3), 205–257.

Wu, Y. (1994). *Mandarin segmental phonology.* (Ph.D. dissertation), University of Toronto, Toronto.

Yang, J.H.-T. (2010). Phonetic evidence for the nasal coda shift in Mandarin. *Taiwan Journal of Linguistics, 8*(1), 29–56.

Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in Standard Chinese: A corpus approach. *Journal of Chinese Linguistics, 42*(1), 218–237.

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