The acquisition of segments and tones in Mandarin: An observational and experimental study

Rui Qi Choo

Doctor of Philosophy

University of York

Education

March 2022

Abstract

This is a study of early word learning and phonological development in children learning Mandarin Chinese in Singapore. Previous work has analysed mostly children's perception and acquisition of segments and tones in the language. The objective of this thesis is to study children's production and examine whether the asymmetry in segmental and tonal information found in *perception* tasks may also be apparent in tasks requiring *production*. Mandarin-learning children's speech forms are systematically investigated here by integrating two strands of research: a naturalistic observational study ($N = 4$) of the influences of long(er)-term knowledge on phonological development in Mandarin is complemented by an experimental study ($N = 20$) of short-term retrieval and production of nonword repetition. The thesis is based on the whole-word approach to the study of children's lexical development and how it may apply to Mandarin, identifying the use of phonological templates and how they may be manifested in Mandarin.

Results from both production tasks reveal independence in the developmental trajectory of segment and tone production. It was not possible to conclusively identify any segmental templates. However, there was evidence of use of two T1-*x* tone templates, which suggest ways in which the whole-word approach might apply here: a salient and well-practiced tone (T1) since the babbling period provides a 'tone envelope' for segments to fill in. The well-practiced T1-*x* motoric routines involve lesser cognitive load, allowing attention to be directed to the (mostly variegated) segmental sequences, so that children may still achieve relatively good matches to the variegated word structures of Mandarin. Thus, there is an interchange between segments and tones and perception and production: tone perception and production begin early but tone production is mastered late, segment perception and production occur later but segment production is mastered before tone is.

Table of contents

List of Tables

List of Figures

Acknowledgements

Thank you, Justin Lo, for being the partner-in-crime for public engagement and social engagement (via online platforms, what/how!?) and the *tai4tai2* to my *tsɛ4tsɛ1* for Cantonese consultations. We probably had a dozen video backgrounds while we shut up and wrote™, including my favourite moment where you spotted instantly the mirrored room in my quarantine hotel. You are in the right field with your sharp eyes (and sharp ears).

Thank you, Margherita Belia, for being the light of my life for research and marking and improv comedy. We will always have the VMS/Qumin joke in our hearts. Thank you, Tamara Bastianello, for your calm companionship. We will always have the afternoon teas, quiet reading and Sunday roasts in our hearts.

Thank you, Sakinah Abdul Aziz, for the decade (and counting) of friendship that has lasted being 22 km apart in Singapore then 300 km via the A1 in the UK then now 10,000 km away. Thank you for being the spot of home away from home, always up for discussing food, and being the constant beam of sunshine and tranquility in the darkness of lockdown days.

Thank you, Pete York, for unintentionally sparking one of the most crucial bits of the thesis: making me realise there are differences amongst the varieties of Mandarin spoken around the world (I had never given them a second thought) via a very memorable *xigua?* comment to Li Sha. Of all the tables in the games café, I chose to sit at the one you were already at (because there were still empty seats); my life in York changed from then.

Thank you, Li Sha and Evelyn Tan, for housing me in my desperate times of need and making things right in the world for those few weeks/months under your respective roofs, for all the picnic joy, games/drinks laughter, blackberry-picking fun and seasons in the sun (and rain and snow) about anything and everything.

Thank you, my second aunt, Xiaoxuan (and sometimes Jingxuan) and Bee Lian for being the north, south and east winds to my west (or any of the directions, all alternated) when I needed a break.

Thank you, my family and J's family, for financial, food and familial support. Thank you, friends in all different parts of the world, whether I knew you in the last year or 10 or 20, for all the wine and cheese we've shared, for the smash-the-patriarchy conversations we've had, for being there through the good and the bad.

Thank you, J – the perennial single-letter acknowledgement which embodies the world, specifically, *my* world. Thank you for being my principal cheerleader, the peaceful presence to my mad mutterings, the provider of everyday necessities and non-necessities, the promise of a future we can finally proceed with now that I've completed this mammoth task that's mostly mine. Now, onwards to our shared milestones ♥

The first degree was the gateway into the working world. The Master's degree was for myself; I found that I liked research and wanted to pursue it further. This PhD is dedicated to the women that came before me: 1) my mum – 谢谢您给我这个机会，让我替您读。
2) both my grandmothers – you did not get a chance to go to school, yet here I am.

I hope I did all of you proud.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1 – Phonological development

This is a study of early word learning and phonological development in children learning Mandarin in Singapore. Given that some phonological templates have been found in children learning Chinese languages, including Cantonese (Mok & Lee, 2018 with 2-year-old children) and Mandarin (Lou, 2020 with 13- to 18-month-old children), the objective of the study is to examine if segmental or tonal patterns may be found in the speech of children aged 20 months and above learning Mandarin in Singapore. Mandarin-learning children's speech forms are systematically investigated in two ways: (i) in a naturalistic everyday setting, from the end of the single word period to the point when they move into combinatorial speech, to establish the templatic patterns and substitutions of segment and tone productions, and (ii) in an experimental investigation into how children perform with regards to segments and tones in Mandarin when faced with stimuli they have never encountered before. Thus, this thesis integrates two research methodologies: a study of the influences of long(er)-term knowledge on phonological development in Mandarin is complemented by a study of the short-term retrieval and production practice involved in nonword repetition.

In this chapter some theoretical approaches to studying language acquisition are presented first, before going on to the basic issue of the thesis: the whole-word approach in the study of children's speech and how it may apply to Mandarin Chinese. Phonological features important to Mandarin will be highlighted in this chapter, with fuller discussion reserved for Chapter 2. An review of the literature on phonological memory follows and the chapter concludes with the focus of the present study.

1.1 Theoretical approaches to language acquisition

How do children represent the language they learn? What children remember of what they hear is influenced by their experience of listening to and speaking particular language(s) with specific features that they need to learn. Exposure to a language entails learning the particular characteristics of the language, which may include segmental and suprasegmental information. Segmental information refers to consonants and vowels whereas suprasegmental features are speech attributes that are not limited to single sounds but extend over syllables, words or phrases (Crystal, 2008). Such features include intonation, stress and speech rate (Sereno & Lee, 2015).

According to Jakobson (1968), there is a universal order to the acquisition of consonants: consonant types that are least marked, that is, that occur frequently in the world's languages, are universally acquired earlier than more marked, less frequently occurring ones. Thus, vowels, plosives and nasals are presumed to be acquired earlier than fricatives, affricates and liquids.

Altvater-Mackensen and Fikkert (2015) offer another perspective as to the order of acquisition of consonants: while principles of markedness (cf. Jakobson, 1968) may be important, articulatory constraints (cf. McAllister Byun, 2012) or perceptual factors (cf. Davis & MacNeilage, 1995) might also influence the acquisition patterns of manner of articulation contrasts. Altvater-Mackensen and Fikkert propose that the distribution of manner of articulation contrasts is influenced by sonority and the contrasts can be defined in terms of syllable structure. For example, less sonorant sounds like plosives or fricatives tend to occur word-initially or in extrasyllabic positions, while more sonorant sounds like nasals or liquids tend to occur word-medially. Since sounds at the beginning of the word generally tend to be more salient to children than those in other parts of the word, plosives or fricatives may thus also be acquired earlier than nasals and liquids as a result. The evidence from babbling concurs with this order: plosives also occur more frequently in babbling than nasals (Gildersleeve-Neumann et al., 2000; Locke, 1983; Vihman et al., 1986).

However, researchers soon found reason to dispute the universal order hypothesis, based on evidence that language characteristics also make a difference in the types of consonants that are acquired first. Studies of child production across different languages have revealed differences in the consonants that children produce first (see Vihman, 2019), depending on the language-specific frequency of occurrence. Children learning English, French and Swedish have been found to produce plosives earlier and more often than fricatives and affricates (classed together in this thesis as 'fricatives'), which appear less often and at a later age (cf. Locke, 1983; Vihman et al., 1994). However, Pye et al. (1987) report earlier acquisition of /tʃ/ in children learning K'iché, despite its low occurrence in the world's languages and its articulatory complexity. A similar story emerged from a comparison study of children learning Xhosa or English: Mowrer and Burger (1991) found that Xhosa-learning children aged between 2 and 6 produced /ts/ and /tʃ/ earlier than their English-learning counterparts. Frequency of occurrence of sounds in a language is thus an important factor to consider. Given that fricatives make up almost half of the consonants that can appear word-initially in Mandarin, fricatives may be acquired earlier than would be predicted by the universal order hypothesis, just like in K'iché and Xhosa. This will be discussed further in Chapter 2.

Meanwhile, work that has emerged since Jakobson has tended less towards universal characteristics and more towards the bottom-up approach of examining overall patterns from children's actual speech data (e.g., Bybee, 2008; Edwards et al., 2011; Pierrehumbert, 2003). From this perspective, every instance of a word that children are exposed to in the language may be learnt as individual lexical items of the same word, known as exemplars. Exemplar models of phonology build on the understanding that the phonological system 'appears to be initiated bottom-up from surface statistics over the speech stream, but refined using type statistics over the lexicon' (Pierrehumbert, 2003, p. 118; also see commentary by Demuth & Johnson, 2020, on learning abstractions from exemplars). This means that a child's phonological system is built up from generalising the patterns from individual instances of lexical items. This would also tie in better with the fact that the patterns that may be observed from children's speech are dependent on the language(s) learnt and the frequency of occurrence of sounds in those language(s).

1.2 Whole-word approach in child phonology

A model that integrates the above perspectives on how children may represent information in their speech is the whole-word approach (see Vihman & Keren-Portnoy, 2013). The key idea is that children do not learn individual sounds; they learn word forms, from which they abstract sound categories (Vihman, 2017), echoing the above-mentioned exemplar models on generalising patterns from children's data.

The whole-word approach stems from work in the 1970s on children's first word production, which found that the child forms were 'holistic' and matched the overall shape of the word better than the individual segments (e.g., Ferguson & Farwell, 1975, Macken, 1979; Waterson, 1971). Menn (1983) offered two options for achieving relative accuracy: children (i) avoid difficult forms (e.g., understanding *b*-initial words but not attempting them at all, cf. Menn, 1976; Schwartz & Leonard, 1982) or (ii) choose to produce their 'favourite' sounds (e.g., actively producing words with fricatives and affricates, cf. Farwell, 1976). These researchers also found that the shape of one part of the word was dependent on another part of the word. This suggests that children are selecting and trying to reproduce the key elements of what they hear in the input. For example, work with a diverse range of languages has shown that some elements of early word forms are less well represented than others, depending on what is salient in their respective languages. Word-initial consonants of trochaic words (words with strong-weak stress) in English are rarely omitted (Vihman & Croft, 2007; Vihman, 2017), because of the salience or the strength of the first syllable relative to the second. On the other hand, children learning languages with word-

final stresses like K'iché produce the stressed second syllable more than the first (Pye, 1992); children learning languages with medial geminates (e.g., Finnish: Kunnari et al., 2001; Vihman & Velleman, 2000; Savinainen-Makkonen, 2007; Italian: Vihman & Majorano, 2017) or consonants lengthened under stress (Welsh: Vihman et al., 2006) also tend to retain the long medial consonants at the expense of the initial consonants. In the latter two instances, the perceptual salience of the stressed syllable and non-initial geminates or lengthened consonants, both of which appear later in the word, draws focus away from the beginning of the words, leading children to omit the word-initial onsets, a pattern seldom found in English phonological development (Vihman & Croft, 2007).

Once children have acquired a sufficient number of words, systematic patterns may begin to be observed. This has been documented in the literature as the first session in which children spontaneously produce 25 or more uniquely identifiable word types in a single half-hour recording session (the 25-word point, or 25wp, Vihman & Miller, 1988). It is found to be the earliest point where a child has built a sufficient sample of words (a cumulative vocabulary of about 50 words) from which generalisations may begin to be abstracted (Vihman & Miller, 1988). These patterns, or 'phonological templates', refer to idiosyncratic child responses to the difficulties of producing segments in different parts of the word and are a stepping stone that helps scaffold children's eventual ability to produce the adult form (Vihman & Croft, 2007; Vihman, 2019). Templates are also an ever-evolving self-(re)organisation and 'summary' of a child's own phonological repertoire. Categories are extracted based on this summary; their repeated use pre-attunes children to the characteristics of the words that they hear in their environment, like an 'articulatory filter' (cf. Vihman, 1993) through which certain sounds 'stand out' to them. According to Vihman (2010), these easily produced and well-practiced templates are motorically accessible (cf. McAllister Byun, 2012), require minimal planning of sequences (because they may simply involve the use of a single consonant or repeated use of the same consonant in a segmental sequence) and may contain familiar segments. Two manifestations of templates may be identified in child forms: words that fit many of the words a child attempts in their target form are termed 'selections', whereas words that children modify to match the template are termed 'adaptations'. For example, Yan Min in the present study had a high percentage of VCV (vowel-consonant-vowel) forms, probably due to her real name having that shape, so her production of *i1fu0* for *i1fu0* 'shirt, or clothes in general' would be deemed a 'selection', whereas her production of *e3te2* for *tɕiɛ3tɕiɛ1* 'older sister' would be an 'adaptation' to the VCV form. In this thesis, templates may thus be identified by high frequency of usage (as a result of both selection and adaptation) alongside moderate precision (i.e., the degree to which a word form is used for targets that match the segment/tone sequence of interest; in

this instance, moderate due to adaptations to the templatic word form). Furthermore, the child's phonological inventory is first considered – mere substitutions for forms that have not yet been acquired are not considered templatic.

There is evidence that segmental templates can be found across various languages (Brazilian Portuguese: Oliveira-Guimarães, 2013; German and Spanish: Kehoe, 2015; French: Wauquier & Yamaguchi, 2013; Finnish: Savinainen-Makkonen, 2007, among others). For example, Brazilian Portuguese-learning Lucas had a template with coda [m] (even though coda [m] does not occur in the language) that was later replaced by a coda with glide [w] (Oliveira-Guimarães, 2013). Kehoe (2015) found that monolingual and bilingual Spanish-speaking children selected more multisyllabic words, selected and produced fewer words with codas, and selected more words with velars (although they did not produce them as such) than their monolingual and bilingual German-speaking counterparts. Additionally, the monolingual Spanish-speaking children also produced more first words containing /l/ than the monolingual German-speaking children. Meanwhile, Savinainen-Makkonen (2007) observed the presence of word-medial geminate consonants in more than 70% of Finnish Joel's speech forms.

As children begin to learn words with increasing complexity and word structures, they need to learn how to tackle *two* onset consonants in a disyllabic word in tandem (C1-C2 sequences, termed consonant sequences; these may also be referred to as 'segmental sequences' in sections discussing the experimental task of nonword repetition as it makes for clearer distinction between segmental sequences and tonal sequences there). This is especially so if they are variegated words (words containing at least two different supraglottal consonants). How templates may present in these specific consonant sequences may be less clear when it comes to manner of articulation, which is the focus of this study. In terms of word structures, substitutions or compensatory mechanisms may first be used by children to accommodate this complexity, either by full reduplication or partial replacement ('consonant harmony') and replacement of consonant(s) by glottals or glides ('other'), or simply, omission of both consonants ('vocalic'). The children's responses may then be examined for these systematic patterns of word structures (e.g., always changing variegated words to harmonised forms; Vihman et al., 2023).

Such templates of word structures have been found in children learning different languages. For example, Priestly's (1977) English-speaking child favoured a specific [CVjVC; C = consonant, V = vowel] template. Consonant harmony has also been found in children learning Italian (Keren-Portnoy et al., 2009) and Brazilian Portuguese (Paulo, in Oliveira-

Guimarães, 2013). Reduplicative and/or CVCV templates have also been identified in children learning Brazilian Portuguese (Oliveira-Guimarães, 2013), French (Wauquier & Yamaguchi, 2013), Arabic (Khattab & Al-Tamimi, 2013) and Finnish (specifically, CVCCV, Savinainen-Makkonen, 2007). VCV patterns have been observed in French: specifically, Béryl (in Wauquier & Yamaguchi, 2013) and Charles (Vihman, 1996) had an <aCo> template (where C = supraglottal consonant. Kehoe (2015) found in her comparison that German-speaking children produced templates with codas while Spanish-speaking children produced more variegation.

Templates have not received much attention in tone languages that contrast suprasegmental (tone) as well as segmental information. There have been some recent attempts at identifying templates in tone languages: Mok and Lee (2018) found evidence of tone templates in children learning Cantonese, whereas Lou (2020) found segmental templates and suggestions of tone templates in children learning Mandarin. Work on perception and production with older children and adults have shown a difference in the way segmental and tonal information is processed and work on *perception* with younger children has also shown this dissociation; more work can be done to examine if this dissociation has roots in development and may also be found in *production* tasks. It remains to be seen whether segmental and tonal templates can be found and the shape they will take. The choice to look at consonant *sequences* and tone *sequences* is two-fold: 1) the majority (70%) of words in Mandarin are disyllabic (Jin, 2011; Zhu, 2002), so it makes sense to study the patterns that occur in disyllables; 2) it came out of the whole-word approach and a decision to look at how the two syllables in a word might interact, or how children learn to produce the entire word structure, in addition to the individual segments, syllables or tones. This is discussed at greater length in Chapter 2.

1.3 Phonological memory

The focus of the previous two sections was on how children draw on the forms available to them in their long-term memory to produce words in their everyday conversations. In this section, the focus is on how this long-term memory both builds on and is itself built by language use and production practice. A review of the literature on phonological memory and nonword repetition is presented next.

Phonological memory, the ability to remember and repeat a novel word form (Vihman, 2022), has been widely studied since the conception of the working memory model (Baddeley, 1986). In a review of work with various English-speaking typically developing adults and

children, as well as neuropsychological populations (e.g., monolingual English-speaking: Gathercole et al., 1997; bilingual Finnish-English-speaking: Service, 1992; Service & Kohonen, 1995), Baddeley and his colleagues (1998) showed that a key function of the phonological loop was to support the acquisition of novel stimuli. They found a link between children's ability to repeat novel sequences of sounds and vocabulary knowledge at a later age. It seemed that the act of repeating and practising the articulation of sounds allowed the sound-meaning connections formed to be stronger. Conversely, children who were not able to produce these temporary novel sequences were less able to create a more permanent representation of the novel sequences in their long-term memory. This repetition of novel stimuli has now been established in the literature as the nonword repetition task.

In the ensuing years, nonword repetition has become a practical screening measure to identify at-risk children in monolingual English-speaking populations, because children with developmental language disorder (DLD; Bishop & Snowling, 2004) are known to be poor at this task (Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990). According to Bowey (1997), not only is nonword repetition a useful measure of phonological memory, it also assesses speech perception (Montgomery, 1995), construction (Snowling et al., 1991) and segmentation (Gathercole & Adams, 1993) of phonological representations. Successful repetition of a nonword is said to involve accurate speech perception, the creation of a temporary but robust phonological representation in working memory, formulation of a motor plan for articulation, followed by the enaction of the motor plan (Coady & Evans, 2008; Snowling et al., 1991).

However, a body of work has emerged that challenges the directionality of the links among nonword repetition, phonological memory and lexical advance. Much of the work relating nonword repetition ability with later lexical advance has been studied in older children, aged 4 years and up, but there is evidence of the existence of phonetic representations before children are able to produce meaningful speech (cf. DePaolis et al, 2011) and without regard to word referents (cf. Yeung & Werker, 2009). Children have been found to be able to abstract patterns from the statistical and distributional properties of the languages they are exposed to in their developing lexicons (cf. Edwards et al., 2011; Metsala & Walley, 1998).

Both perception and production research provide evidence of how children abstract patterns from the words they learn, rather than children learning individual segments and piecing them together to learn words. In a review of the perception literature, Coady and Evans (2008) state that children store words in their lexicons not in sequences of phonetic segments, but instead as holistic and syllable-based units (e.g., Jusczyk, 1986; Walley,

1993). This is in line with whole-word accounts of the production patterns in children's early words (see section 1.2, above): children's first words tend to be holistic and more accurate than expected (i.e., they resemble the adult form: cf. Ferguson & Farwell, 1975). As discussed above, target language prosody has been shown to influence the production of word forms as a whole; some elements of early word forms have been shown to be less well represented than others (e.g., unaccented syllables and words with medial geminates).

Evidence that phonological memory both builds on and is itself built by language use and production practice (e.g., DePaolis et al., 2011; Keren-Portnoy et al., 2010; Zamuner & Thiessen, 2018) may be shown by the developmental changes in the bidirectional process of existing phonological and lexical knowledge guiding future lexical development. At an earlier age (that roughly coincides with a less vibrant lexicon), accurate production of newly encountered words has been shown to depend on prior knowledge. Zamuner and Thiessen (2018) found that a significant predictor of whether a child between the ages of 1 and 2 years imitated a new word was whether the word contained more sounds that the child had correctly produced before. This concurs with Keren-Portnoy et al. (2010), who tested 26-month-olds' nonword repetition ability and found that nonwords with consonants that were already in the child's repertoire (IN consonants) were better produced than those with OUT consonants (see also Leonard et al., 1981). These results are consistent with the 'articulatory filter' hypothesis (originally proposed for children at the one-word stage but still relevant here; Vihman, 1993; Vihman et al., 2014), which states that prior practice with known sounds provides a filter through which subsequent words that contain these familiar sounds appear to 'stand out' to children. Words containing familiar sounds are facilitated by the children's previous production experience, which in turn helps in the retention and construction and learning of future word forms with those sounds (McAllister Byun & Tessier, 2016; Vihman, 2022; Zamuner & Thiessen, 2018).

Two types of tasks have been shown to elicit children's preference for and learning of unfamiliar items over familiar ones. DePaolis and colleagues (2011, 2013) found in their studies of 9 to 11-month-old infants that those who consistently produce only one consonant preferred listening to passages containing nonwords with that consonant, whereas infants who were consistently producing more than one consonant (i.e., who were more phonologically advanced) preferred listening to passages containing nonwords with consonants *unfamiliar* to them. Storkel's (2005; 2006) work with children aged 3 to 5 years (a group of children with phonological delays, age-matched children and vocabulary-matched children) showed that regardless of typical development or phonological delay, children with a more established expressive lexicon (well beyond the 25-word point, refer to

Chapter 2) learnt items containing OUT sounds better than those with IN sounds. These findings show that children who have established a solid repertoire of sounds or words seek novelty; older children also require fewer resources in the learning of what is familiar and instead devote resources to the learning of what is as yet unfamiliar.

Phonological memory is thus constantly supported by and itself supports passive learning and active processing of the ambient language and babbling practice (Keren-Portnoy et al., 2010; Vihman, 2022). In a similar vein, López Assef et al. (2021) describe a developmental shift in the production effect (i.e., where production of a stimulus facilitates recall) where younger children (aged 2 to 3 years) have been found to recall a nonword better when they have simply *heard* a word than told to produce it but older children (aged 5 to 6 years) do better at recall when told to *produce* the nonword.

As mentioned, researchers can test what stays in children's memory when they learn words by means of the nonword repetition task. A discussion of nonword repetition follows.

1.3.1 Nonword repetition

Phonological memory is often operationalised using the nonword repetition task. Early proponents of the task claimed that nonword repetition is a relatively 'pure' measure of phonological memory, because children cannot rely on long-term knowledge when nonwords are used (cf. Gathercole et al., 1991; Gathercole, 2006); Ellis Weismer et al. (2000) also claimed that it is a 'content-free' measure, which makes it culturally appropriate to use across various populations. However, some authors have countered the notion that it is a 'pure' measure of phonological capacity because long-term memory information has been shown to influence short-term memory performance (cf. Jones & Macken, 2018). For instance, even Gathercole and colleagues' early papers (e.g., Gathercole et al., 1991) found a 'wordlikeness' effect (the degree to which a nonword resembles a real word), i.e., the closer the nonword conformed to the phonological structures of the language, the better the children's abilities to repeat those nonwords. As mentioned in the previous section, children abstract and derive knowledge of the phonological structure of their language from their own vocabularies (Bowey, 1997), and this ability to recognise systematic patterns in the language assists them in their retention and repetition of nonwords that contain familiar sequences (Dollaghan et al., 1995; Keren-Portnoy et al., 2010; Snowling et al., 1991). Stress patterns also make a difference: the production of multisyllabic stressed and non-stressed syllables is better when they match the stress patterns of real words (e.g., Dollaghan et al., 1993; 1995);

whole syllable errors in nonword repetition tasks involved omission of unstressed syllables (e.g., Roy & Chiat, 2004).

Researchers who have statistically operationalised 'wordlikeness' have found better performance for nonwords with high than low wordlikeness, as mentioned. For instance, nonwords with high phonotactic probability (the frequency with which segments and sequences of segments occur in words in any language, sometimes calculated using biphone frequency), have been found to be more accurately produced than words with low phonotactic probability in adults (e.g., Vitevitch & Luce, 2004, 2005) and children (e.g., Beckman & Edwards, 2000; Edwards et al., 2004; Hoff et al., 2008; McAllister Byun & Tessier, 2016; Zamuner et al., 2004).

Other authors found a corresponding decreasing order of accuracy for repetition of words vs. nonwords (e.g., Roy & Chiat, 2004; Keren-Portnoy et al., 2010), and nonwords with high vs. low phonotactic probabilities (e.g., Gathercole et al., 1999). Individual features of words have also been shown to make a difference: better performance has been recorded for nonwords with features that are within rather than outside a child's own phonetic inventory (e.g., Keren-Portnoy et al., 2010; Stokes et al., 2006). Nonwords have also been shown to prime the semantic targets of real words with which they share acoustic-phonetic features (e.g., Connine et al., 1993). Additionally, Dollaghan et al. (1995) found a lexicalisation effect, with nonwords being produced as their phonological real word neighbours, reducing nonword repetition accuracy. The authors caution that nonword repetition stimuli need to be checked for their similarities to actual words not merely in terms of whole words, but also on a within-syllable basis. The phonotactic rules of the language as well as the frequency of particular phonotactic probabilities of consecutive sequences must thus be taken into account. These accounts also bring up the importance of controlling for frequency effects. According to Ambridge et al. (2015), frequency has always been considered a 'ubiquitous' factor in language acquisition. In developmental research, the relationships between word frequency and production accuracy remain unclear, mostly as different frequency outcome measures have led to different results (Stoel-Gammon, 2011). Curiously, some researchers have used adult corpora even on studies with 2- to 3-year-old children (e.g., using Kučera & Francis, 1967: Sosa, 2008; Sosa & Stoel-Gammon, 2012), which is surprising, as the frequency counts for adult words is likely different from child-directed words, with corresponding differences in familiarity to the children. A better practice would be to use child corpora as a basis on which to study child productions (e.g., the use of the CHILDES database, MacWhinney, 2000 by Tang et al., 2019a, b). Indeed, the familiarity of words has been found to have an effect on accuracy of production: children learning Cantonese were significantly

less accurate when producing less familiar words than highly familiar ones, whereas there was no such difference for adults (e.g., Wong et al., 2017).

Others have used frequency counts based on the child's own word production (e.g., in English, counting how often a child uses a word: Tyler & Edwards, 1993), which is also not ideal, as counting child usage of words may be idiosyncratic and may not be generalisable to other children [e.g., previous examples about a child favouring CVjVC words or adapting them to this pattern (Priestly, 1977) or a child actively choosing words with fricatives and affricates (Farwell, 1976)] or indirectly favour words with certain articulatory properties. For example, 'car' may occur about ten times as often as 'chair' (cf. Vitevitch & Luce, 2004[1]), because the /k/ in *car* is less challenging for children to produce and perceive than the /tʃ/ in *chair*, even though both items may occur readily in the child's environment (and further confounded by the fact that a car may move and may be more exciting to a child than a typically-stationary chair may be). The articulatory properties of words thus need to be taken into account as well. In this thesis, since the focus is on disyllabic words in Mandarin, it makes sense to control for the frequency of consonant *sequences* and the corresponding articulatory properties of the two onset consonants in tandem (onset segments and not codas are studied since there are only a limited number of codas [n and ŋ] in Mandarin). Consonant sequences in the language are first measured and ranked in order to determine the frequency of occurrence of respective consonant sequences. This ties in to the study of nonword repetition in Mandarin.

1.3.2 Nonword repetition in Mandarin

As presented, statistical phonological structures in a language influence the repetition of stimuli, even if the stimuli were *non*words. If the nonword repetition measure were truly 'pure' (cf. Gathercole, 2006), there should have been no difference in performance on the above counts. The evidence presented so far has also only been based on English. Researchers studying children who speak other languages need to create nonwords that resemble the specific language features of that language. For a tone language like Mandarin, nonwords in Mandarin need to account for both segmental and tonal aspects of the language. Frequency effects must also be considered (e.g., wordlikeness, phonotactic probability), and for Mandarin, this includes calculating the frequency of occurrence not only of sequences of

---

[1] In the spirit of using child-based corpora, the Child Phonotactic Probability Calculator (Storkel & Hoover, 2010) was sought. Unfortunately, it was inaccessible even at the point of thesis submission and resubmission (see explanation from the website developer: http://www.people.ku.edu/~mvitevit/PhonoProbHome.html), hence the adult Phonotactic Probability Calculator was used.

segments but also of sequences of *tones*. For example, Parra et al. (2011; see also Jones et al., 2004) tested children on English-like and Spanish-like nonword stimuli and found both language-general and language-specific components to phonological memory. There was a correlation between the phonological memory for English-like and Spanish-like nonword stimuli, as well as language-specific benefits of language exposure and of phonological memory to language development. Service (1992; Service & Kohonen, 1995) found that Finnish-learning children's performance on an English-like nonword repetition task predicted their English learning. The fact that language-specific effects can be found shows that the task is not entirely 'content-free' (cf. Ellis Weismer et al., 2000), and has to be designed to fit the particular language.

Only a few groups of researchers thus far have studied nonword repetition in Chinese languages and these have all been geared towards atypical language development. For example, Ho and Lai (1999) studied naming speed deficits and phonological memory deficits in Chinese children with developmental dyslexia in Hong Kong. For the phonological memory tasks, children were asked to repeat nonsense Chinese (presumably, Cantonese) syllables, but the authors did not provide examples of these syllables. It is likely that it is a monosyllabic Cantonese nonword repetition task to compare with their other measure of digit span (digits from 0 to 9 are monosyllabic in Cantonese), but it is nonetheless unclear whether the syllables themselves were nonwords or if the syllable was a legal word in another tone but the tone implemented for that syllable in their task made it a nonsense syllable (termed a 'tonotactic gap', see later reference in section 3.2.1). Additionally, since their study was aimed at testing the naming-speed deficit hypothesis on children with developmental dyslexia learning non-alphabetic scripts, the element of segment and tone in their nonword repetition task was not their main concern.

To the best of the investigator's knowledge, Wang and Huang (2016) were the only ones who studied Mandarin nonword repetition (also aimed at diagnosing atypical language development) who included the nonword repetition stimuli used. The authors administered a word and nonword repetition task to Mandarin-speaking preschool children suspected of having developmental language disorder (DLD) and age-matched controls (23 children per group, aged between 3 and 6 years). The list of five words included [forms provided are as used in their study with tone added – the authors did not provide a tone for the real words]: *pai2tʰian1* 'day', *nan2xai2* 'boy', *lau3tʰəu0* 'old man' (colloquial), *njəu2nai3* 'cow's milk' and *xau3kʰan4* 'nice-looking'. Nonwords were then created either by metathesising the vowels, changing the initial consonants, or completely replacing entire syllables of these real words [forms provided are as used in their study; all the nonwords were said in T1-1]: *pia tian*, *nua*

fai, *luai tʰiu*, *miəu nuai* and *xiɔ kʰian*. They found that the former group of children had no issues with tone in the nonwords but nonetheless performed worse than age-matched controls on nonword repetition. The authors thus concluded and agreed with Ellis Weismer et al. (2000) that the nonword repetition task was 'bias-free', and that two-syllable Mandarin nonwords can be used to diagnose Mandarin language disorders. Their conclusion is contentious for a few reasons. First, they used only five real words and five nonwords in their task, which seems too small a sample of words for a conclusive finding. In terms of their stimuli, all the nonwords were said in the same tone (T1, high level) in order to prevent children from being distracted by homophonous words as well as to have consistency across the experimenters who had to administer the task 'live'. However, T1 tends to show the earliest acquisition and highest levels of precision (i.e., the degree of match of child form to target) as it is the most salient tone in Mandarin (see further discussion in Chapter 2), so the lack of tone errors in the nonword repetition task is not all that surprising. Next, their 'bias-free' claim was based *solely* on the fact that the children suspected of DLD and age-matched controls scored significantly differently on the nonword task, and not based on something other than their performance on the very same nonword task that they were administering, which seems to be a circular argument. Finally, as a result of using the same T1 for all nonwords, how tone influences nonword repetition could not be studied.

1.4 Present study: What segmental and tonal patterns occur in children's Mandarin speech productions?

Most of the research on the dissociation between segment and tonal processing in Mandarin has covered adult and child perception. Less is known about the potential joint effects of segments and tone in development, especially for *production*. More work should be dedicated to examining the (a)symmetry in the way that segments and tone influence language production, especially from a developmental perspective. What are the salient linguistic features that children pick up on and what patterns do they produce when they are beginning to speak? Work with older children (toddlers vs. preschoolers, cf. Singh et al., 2015) has found a shift in relative sensitivity to segments vs. tone information, which means it is necessary to study children from an earlier developmental timepoint, when they are just beginning to speak Mandarin, to try to establish the origin of the segment/tone asymmetry.

From a psycholinguistic viewpoint, production studies are necessary to understand how children process what they hear and how they represent it as speech. In addition, production involves additional, serial or simultaneous processes of articulatory motor control, cognitive planning and actual production (Keren-Portnoy et al., 2010). The early infant perception

literature showed how children quickly attune to the distributional frequencies of units of the language(s) they are exposed to (e.g., Burns et al., 2007; Eimas, 1978; Kuhl et al., 1992; Mattock & Burnham, 2006; Werker & Tees, 1984). How children learning Mandarin are influenced by the frequencies of segment and tone information in the language in their production of disyllables has not yet been systematically studied. While previous studies on child perception have examined the three-way distinction between consonants, vowels and tone, work on child production has shown that it is difficult for transcribers to transcribe reliably the vowels produced by children at the end of the single word period (cf. Davis & MacNeilage, 1995; Davis et al., 2002; Keren-Portnoy et al., 2010; Vihman, 2019). Thus, for this thesis, the present analyses are restricted to the consonant/tone asymmetry. Henceforth, the term 'segment' refers to 'consonant' unless otherwise stated.

For Mandarin, the accuracy of a word depends on both segment and tone accuracy. If either unit is incorrect, e.g., an inaccurately produced segment with an accurately produced tone or an inaccurately produced tone with an accurately produced segment, the entire word is considered incorrect. For disyllabic words, this issue is compounded: both the segments and the tones of each syllable need to be accurately produced in order that the entire disyllabic word be said accurately.

Error analyses have been useful in examining the elements of words that are represented when a child learns languages. For beginning learners, speech errors in Mandarin are likely to occur in either one or both segment and tone elements. In adult slips-of-the-tongue errors, the rate of co-occurrence of errors between segments and tones have been found to differ: tone errors have been said to occur less often than segmental errors (Chen, J.-Y., 1999; 2000): tone tends to remain intact even if there are within-word segmental errors, i.e., only if the errors are entirely segmentally incorrect would tone *also* be incorrect. The (a)symmetry of segmental and tonal information in Mandarin is discussed in greater detail in Chapter 2.

Errors in child forms may also tell us the important elements in phonological encoding (cf. O'Seaghda et al., 2013), and how words in a language may be represented holistically (cf. the whole-word approach). As mentioned above (section 1.2), children learning languages with certain prominent features may make more errors relating to the *less* prominent features. Thus, these error patterns allow us to make inferences on how language may be represented in the mind of the learner, dependent on the salient features of the words in the respective languages, which differ in their location between one language and another. The challenge in Mandarin is then whether words in Mandarin may be represented holistically and the type of error patterns that may be found, in both everyday speech and a

nonword repetition task. Tone errors have been found to occur more frequently than segmental errors in adult word production (Chen, J.-Y., 1999; 2000). Which element, then, segment or tone, would children be more likely to produce accurately in everyday conversation and grasp and subsequently repeat accurately in a nonword repetition task? The fewer tone than segmental errors found might indicate that tone is a more salient feature than segments, possibly because it is a suprasegmental element that children have been paying attention to from a young age (cf. Fernald, 1989). Fewer tone than segmental errors may then also be expected in both production tasks. Additionally, for the nonword repetition task, given the dissociation found in previous work, it is also not clear if there would be an influence on the frequency of segment and tone sequences on segment and tone accuracy and a crossover effect, i.e., whether frequency of segmental sequences would influence tone accuracy and vice versa.

The objective of the present study was to examine the systematic patterns that may be evident in children's segment and tone productions using two complementary approaches. The first is a longitudinal observation that tracks children's productions in a naturalistic play setting with a parent over six monthly recordings. The purpose of the naturalistic observations was to examine children's productions of segmental and tonal information in their everyday speech. The manner categories of word-initial consonants as well as the word structures of the words produced were analysed, and patterns of production and substitution and any uses of templates were noted. Manner categories were the focus of this study because the frequency of occurrence of fricatives (and affricates) in the consonant repertoire in Mandarin (about half) might lead to their (early) order of acquisition, as has been observed in some of the languages discussed above (in section 1.1; Altvater-Mackensen & Fikkert, 2015). The term 'word structures' as used in this study is used for 'prosodic structures' in other studies (e.g., Lleó, 2002; Pierrehumbert, 2003; Vihman, 2019 etc.); however, since prosody may refer to suprasegmental elements of speech like tone, stress and intonation (Crystal, 2008; Sereno & Lee, 2015), and with tone being an important factor in this present study, the term 'word structures' is adopted instead to prevent confusion. This longitudinal observation also provides an extension of the findings by Lou (2020) on 13- to 18-month-olds; comparison and reference to Lou's study will be made where relevant. The present study was carried out with four children in Singapore, beginning when they were 20 months old. The study of children at a slightly later lexical stage provides a continuation and exploration of the kinds of patterns that children may produce early in the word-combination period.

The second is the administration of an experimental nonword repetition task created for this study. The purpose of the nonword repetition task was to examine the accuracy of segment vs. tone production in children's productions of unfamiliar sequences. If segment and tone templates may be observed in children's everyday speech, children might also use templates in their responses to unfamiliar stimuli. Like the naturalistic observations, the manner categories of word-initial consonants as well as the word structures of the words produced were analysed, and patterns of production and substitution and uses of templates noted.

In addition, the effects of frequency were also controlled for: whether and how the frequencies of occurrence of segmental sequences (segmental feature) and tone sequences (suprasegmental feature) influenced the segment and tone accuracy of the nonword repetition task were analysed, as well as whether there was an interaction between the two factors. Wang and Huang's (2016) nonword stimuli were not controlled for frequency, which could have been a confound. Both frequency of segmental sequences and *tone* sequences should be controlled for when creating novel Mandarin stimuli. Furthermore, child corpora should also be used to calculate the frequencies when the participants of study are children [cf. the use of the CHILDES database by Tang et al., 2019a, b). Since the study was administered to Singapore children, the Singapore Communicative Development Inventory (CDI; Singh et al., 2022) was used to measure the frequency of words that children in Singapore learn. The CDI had been adapted to suit the Singapore population: for example, when adapting the words from the original MacArthur-Bates CDI (Fenson et al., 1993), words like 'snow' and 'lawnmower' were removed as Singapore is in a temperate climate with no snow, and lawnmowers are not common in the Singapore context where most people live in apartment blocks with no lawns. Within the adaptation, the more familiar form of lexis to Singapore children was chosen for certain words (e.g., 'potato' in Putonghua is *tʰu3tou4* but *ma3liŋ2ʂu3* in Singapore [and Malaysia and Taiwan] Mandarin), in line with word familiarity effects found in other studies (e.g., Wong et al., 2017). The CDI served not only as a measure of words children say but also the words they understand, so it is a good proxy for input to the children as well as for child output.

Thus, the invented words were disyllabic, conformed to Mandarin word patterns and took into consideration the frequencies of individual segmental sequence and tone sequence combinations as they occurred in the language. However, whether there may be use of the same templates, or any use at all in the nonword repetition task, is unclear, due to the internal segment and tone frequencies of the task, which differs slightly from frequencies in the language (full details of the creation of the nonword repetition stimuli are presented in

section 3.2.1). The repetition task also included familiar words to make the task less daunting for the children (but only the nonwords were analysed). This study was carried out with 23 2-year-old children in Singapore (of which 20 produced enough responses to be included in the analyses).

The key questions for both naturalistic and experimental portions of the study are:

1) Is there overlap among the consonant sequences and tone sequences children hear, target and produce?
2) Is there evidence of templatic usage in the children's segment and tone productions?

An additional question is asked of the experimental study:

3) How, if at all, does the frequency of occurrence of segmental sequences and tone sequences influence the accuracy of child productions of newly encountered stimuli?

This chapter has covered theoretical approaches to language acquisition, focusing on the whole-word approach on which this study is based, and has reviewed the literature on phonological memory and how a nonword repetition task in Mandarin may be administered. Chapter 2 covers the literature on phonology of segments and tones, and the phonological acquisition of segments and tones in Mandarin Chinese, including the identification of templates in Mandarin. There is also a section on Mandarin as it is used in Singapore, followed by a review of findings regarding the dissociation of segmental and tonal features in Mandarin perception and production studies. Chapter 3 covers the methodology of the thesis, elaborating on how the consonant sequences and tone sequences were transcribed and analysed and how the frequencies of occurrence of consonant sequences and tone sequences of the language were taken into account in the design of the nonword stimuli. Chapter 4 discusses the findings of the longitudinal, naturalistic observation study in terms of segmental analyses (section 4.1) and tonal analyses (section 4.2), focusing on the production patterns of consonant sequences and tone sequences, their substitution patterns and the identification of any templatic usage. Chapter 5 discusses the findings of the experimental nonword repetition study, including the same segmental and tonal analyses as in Chapter 4, with additional statistical analyses to support the descriptive findings. The final chapter rounds up the over-arching research question addressed in this study – what do the findings from children's segment and tone productions and patterns tell us about phonological development in Mandarin?

Chapter 2 – Phonology and phonological acquisition of Mandarin Chinese

Despite the fact that more than half of the world's languages are tone languages (Yip, 2002), the literature has disproportionately focused on non-tone Indo-European languages which contrast words via changes to segmental elements like consonants and vowels only and do not have suprasegmental features like tone that are of lexical significance. The situation is more complicated for tone languages, where contrasting the lexical pitch of a word makes a distinction in meaning.

2.1 Phonology of Mandarin

Mandarin is the most widely-spoken tone language in the world, where word-level pitch variations are used to differentiate meanings, either at the syllable- and/or word-levels of linguistic representation (Liu et al., 2018; Maddieson, 2011; Yip, 2002). Mandarin belongs to the family of Sino-Tibetan languages (Katzner, 1977). It is spoken by 885 million people, 87% of whom live in China; most of the rest in Taiwan, Singapore, Malaysia and Indonesia (Gordon, 2005). Mandarin is one of the six official languages of the United Nations; it has official language status in China, Taiwan, Hong Kong, Macau and Singapore (Erbaugh, 1992; Wee, 2013). The Beijing variety is taken to be the standard in China, known as 普通话 *Putonghua* 'Common Language' (IPA: *pʰu3toŋ1xua4*). It is also known as Putonghua in Hong Kong and Macau, but 国语 *Guoyu* 'National Language' (*kuo2y3*) in Taiwan and 华语 *Huayu* 'Chinese Language' (*xua2y3*) in Singapore and Malaysia (Chen, P., 1999).

2.1.1 Segments

Words in Mandarin are made up of syllables and tones. When referring to tone languages, 'syllables' refer only to the segmental properties of the language without tone, i.e., an atonal syllable; syllables with tone will be explicitly referred to as such. Syllables in Mandarin are simple, with single-consonant-onsets (no consonant clusters) and end in a vowel and/or nasal /n/ or /ŋ/ coda. There are clear syllable boundaries in Mandarin, with little resyllabification and no ambisyllabicity (Duanmu, 2000). The consonants do not carry prosodic information; they are realised using aspiration, manner and place of articulation; vowels, on the other hand, carry prosodic information, stress, intonation and tone (Burnham et al., 2011; Lin et al., 2013). Tones and the vowel are compulsory in a Mandarin syllable, whereas the onset and coda are optional (Zhu & Dodd, 2000). In Mandarin, there are 21

consonants that can occur as onsets: 6 plosives (28%: p, pʰ, t, tʰ, k, kʰ), 11 fricatives and affricates (classed together as 'fricatives', 52%: f, tɕ, tɕʰ, ɕ, ts, tsʰ, s, tʂ, tʂʰ, ʂ, h), 2 nasals (10%: m, n) and 2 liquids (10%: l, ɹ). There is also an additional consonant [ŋ] that can only occur in a syllable-final position. Only /n/ and /ŋ/ can appear in the coda. In this thesis, reference to 'segments' refers to '(onset) consonants', unless otherwise specified.

There is no consensus regarding the syllable structures of Mandarin. Some researchers have argued that there are no diphthongs and triphthongs in the language; instead, the underlying representation of a rhyme in Mandarin is a glide + vowel + optional coda combination. This is because the glide is said to retain some unique phonetic and phonological qualities (Wan & Jaeger, 1998). However, as an in-depth discussion would take us too far afield here, there will be only brief mention of the issue where relevant (see discussion in Wan & Jaeger, 1998). For the purposes of this study, since Zhu's (2002) study has been an influential account of Mandarin phonology, her notation of vowels is followed, which includes monophthongs, diphthongs and triphthongs. Following Zhu (2002), there are 22 vowels in Mandarin: 9 simple vowels (i ə ɚ ɤ ɛ y o a u), 9 diphthongs (ai au ei ou ia iɛ ua uo yɛ) and 4 triphthongs (iau iou uai uei). There are three allophones of /i/: it is produced as /ɿ/ after /ts/ /tsʰ/ and /s/; /ʅ/ after /tʂ/, /tʂʰ/, /ʂ/ and /ɹ/; and /i/ after all other consonants.

Mandarin words are constructed from a relatively small inventory of syllable units (Packard, 2000). There are 420 possible combinations of onsets and rimes in Mandarin and 1,300 syllables with tone if tonal variations are accounted for (*Xiandai Hanyu Cidian* 'Modern Chinese Dictionary', 1979; cited in Zhu, 2002; Chen et al., 2002; Deng & Dang, 2007; Lin & Wang, 1992; O'Seaghdha et al., 2010). As a result, homophones occur much more frequently in Mandarin than in other languages (Zhu, 2002). Ke (2006) compared the homophony in 20 Chinese languages and 3 Germanic languages. Using the Dictionary on Computer electronic database of the phonological systems of Chinese languages (Wang, 1969), Ke found 67% homophony in the Beijing dialect (which is the dialect referred to as 'Mandarin') from over 4000 entries. Using corpus frequency from the CELEX database (Baayen et al., 1995), among the first 2000 most frequent words, Ke (2006) found about 10% homophony in English and about 4% each in Dutch and German.

Mandarin is said to be morphosyllabic (DeFrancis, 1984): each syllable is a morpheme in itself and words can be made up of two or more morphemes. Most words in Mandarin are disyllabic (74%; Institute of Language Teaching and Research, 1986, cited in Zhou & Marslen-Wilson, 1995) or made up of more than two syllables (Zhu, 2002). Jin (2011) provides an account of the homophone avoidance approach (cf. Guo, 1938; Karlgren, 1949;

Li & Thompson, 1981; Lü,1963) which explains why there is a large percentage of disyllabic words in Mandarin (70% in their study) compared with American English (38%), Cantonese (30%), Hawaiian (28%) and Japanese (17%; trisyllabic words are most prevalent in Hawaiian and Japanese). The approach states that Chinese monosyllabic words were lengthened to avoid semantic ambiguities that would have arisen as a result of homophony. Based on these features, the syllable seems to be an ideal 'proximate unit' (cf. O'Seaghdha & Chen, 2009) for lexical processing in Mandarin.

2.1.2 Tone

There are four tones in Mandarin, Tone 1 (T1): high level (55 in Chao, 1930 notation), T2: rising (35), T3: falling-then-rising (214, or only falling [21] in some tone combinations, see below), T4: high falling (51). Tones are associated with individual syllables (Yip, 1995), are carried by the vowels (Lin et al., 2013), and the function of tone is predominantly lexical (Wan & Jaeger, 1998): each syllable has a different meaning depending on the lexical tone. The syllable *ma* can refer to 'mother' in T1, 'hemp' in T2, 'horse' in T3 and 'scold' in T4. The final rise of T3 is said to occur only in isolation, word-finally or at the end of a sentence (Chen, C. Y., 1984; Singh & Fu, 2016; Tang et al., 2019b). The lower pitch 21 appears elsewhere (also termed the 'half-T3 sandhi': Tang et al., 2019b; Zhang & Lai, 2010), under normal stress, in syllables preceding T1, T2, T4 and T0 (Chen, C. Y., 1984; Wang & Li, 1967). Just as consonant and vowel changes result in lexical contrasts, so too do tone changes (cf. Burnham & Mattock, 2007).

Most words in Mandarin have a tone in which they are consistently produced, but sometimes the morphological structure of Chinese words dictates that some words undergo tone sandhi, the alternation of tones. There are four lexical items with their own sandhi rules: *pu4* 'no', *i1* 'one', *tɕʰi1* 'seven' and *pa1* 'eight'. They are produced as a rising tone (T2) before a high falling tone (T4). Additionally, *i1* is said as a high falling tone (T4) before all other tones except a high falling tone (T4).

The sandhi rule that has garnered the most attention is the T3 sandhi rule. It states that when there are two consecutive T3 syllables (T3-3), the first of those syllables will become a rising tone (T2), with the resulting tone sequence being T2-3. Hockett (1947) stated that this substituted rising tone is at a lower pitch than that of a true T2. However, Wang and Li's (1967) tone perception experiment with adults showed that Mandarin speakers could not differentiate between the pitch of a true T2 and one substituted via the T3 sandhi rules. Ching (1971; cited in Chen, C. Y., 1984) further rationalised two instances of when the

sandhi rule occurs: first, the T3 sandhi rule change in the first syllable only applies when the second syllable is a "substantive morpheme or a 'functional reduplication'" (p. 303). For example, the verb 'walk' may simply be represented by the lone morpheme *tsou3*, but when it is reduplicated it takes on a functional meaning of 'take a walk'. *tsou3* reduplicated is *tsou3tsou3*, and after sandhi rules apply the eventual form is *tsou2tsou3*. Second, the T3 sandhi rule change does not happen when it is an "empty morpheme or a 'simple reduplication'" (p. 303). For example, 'older sister' may simply be represented by the lone morpheme *tɕiɛ3*; when it is reduplicated there is no functional change. *tɕiɛ3* reduplicated is *tɕiɛ3tɕiɛ3*, but as it is a simple reduplication, sandhi rules do not apply: rather than *tɕiɛ2tɕiɛ3*, the resultant form has a neutral tone *tɕiɛ3tɕiɛ0* instead.

While researchers are unanimous in agreement about the existence and delineation of the pitch contours of the four tones in Mandarin, whether there is a fifth tone (Tone 0, T0) is more contentious. Researchers are mixed as to whether the fifth tone should be considered a 'neutral tone' (Norman, 1988) or simply 'weak stress' (Zhu, 2002), and whether these two are the same. According to Chen (1984), every stressed syllable has a full tone, and when a syllable is weakly stressed (Chao, 1968), its duration becomes shorter and its tonal characteristics are neutralised, hence the term 'neutral tone'.

According to Norman (1988) and Zhu (2002), neutral tone or weakly stressed syllables cannot occur in isolation, unlike the four basic tones. Furthermore, the neutral tone does not have a consistent pitch contour: its pitch is determined by the preceding tone. It is half-low after high level T1, mid after rising T2 and low after high falling T4 (see Zhu, 2002 and Norman, 1988 for more detailed analyses; only the most relevant points are discussed below). For falling-then-rising T3, when such a syllable precedes the neutral tone, the final rise (the **4** of the 214 tone contour) in pitch is realised on the syllable of the neutral tone (Wang & Li, 1967). Presumably, the neutral tone has mostly a mid-level pitch to it that is between the high level T1 and high falling T4. The weak articulatory strength of the neutral tone is said to account for its variability (Chen & Xu, 2006).

Zhu (2002) posits that weak stress or non-stress and the neutral tone are linked. Hockett (1947) and Stimson (1967) caution against a simple dissociation between weak stress and the neutral tone. They state that syllables with weak stress and the neutral tone do not always interact as cause and effect; there are weak-stressed toned syllables as well as normal-stressed toneless syllables. Zhu (2002) further elaborated on stress in Mandarin syllables, differentiating between words with regular and irregular weak stress. The former includes weak stress which typically occurs in functional words and morphemes like particles

(e.g., sentence-final grammatical particle *lə0* denotes something that's already been done), affixes (e.g., *tsɿ* in *pi2tsɿ0* 'nose') or on the second part of reduplicated nouns, verbs or adjectives (e.g., *ɕiŋ1ɕiŋ0* 'star', *ɕiɛ4ɕiɛ0* 'thank you') (Chen, J.-Y., 1999; Zhu, 2002). Sometimes weak stress serves to differentiate words in contrast to words with normal stress or a full tone (e.g., *toŋ1ɕi0* 'thing' vs. *toŋ1ɕi1* 'east-west'; Chen, J.-Y., 1999). These are not frequent in the language and the context usually helps with disambiguation. Words with irregular weak stress are not rule-governed and whether a syllable is weakly stressed or not is decided on a lexical basis, referred to as 'lexeme' type, i.e., the second of two lexemes is weakly stressed when they are combined together, e.g., *ɚ3tuo0* 'ear', *tʰou2fa0* 'hair' (Zhu, 2002).

The use of the neutral tone is even more sporadic when it comes to the nomenclature for family members. Chen (1984) compared kinship terms in the Modern Chinese Dictionary (*Xiandai Hanyu Cidian*) and found inconsistencies relating to the usage of the neutral tone. For example, comparing the names for men (*fu*, in **bold**) and women (*mu*, underlined) on the maternal side[2]:

> *i2**fu0*** 'aunt's husband' vs. *i2<u>mu3</u>* 'aunt';
> *tɕiou4**fu4*** 'uncle' vs. *tɕiou4<u>mu0</u>* 'uncle's wife'

Chen (1984) concluded that there were no clear consistent cues that can be found in the grammatically unpredictable instances of the neutral tones. However, the investigator proposes the following pattern, one that distinguishes between an aunt/uncle who is or is not related by blood: a full tone is attributed to the *fu* and *mu* when it is the parent's *direct* sibling, whereas the neutral tone is attributed when it is the *spouse* of a parent's sibling.

In the present study, kinship terms will turn out to belong to a collection of tone templates in themselves in Singapore Mandarin, including one with a neutral tone. Kinship terms are familiar to children and emerge early in child production (Zhu, 2002), hence it is unsurprising to find templatic patterns in this group of words. These are further discussed in Chapter 4.

---

[2] The naming convention that many Chinese families adopt is that the person's actual name is not used, rather, the relationship is emphasised. For instance, there may be a number (denoting birth order), a word that represents the sex (aunt or uncle), lineage (maternal or paternal side) and spouse (husband or wife of said aunt/uncle). Thus, via this honorific, one can immediately work out the familial relations. This is in contrast to other cultures which teach children to call the siblings of their parents Aunt Mary and Uncle John, or simply, Mary and John.

Liu and Samuel (2007) conclude that tones are more important than syllables or segments because of the amount of "segmental homophony" (Liu & Samuel, 2007, p. 589) that results from the simple syllable structure (mostly consonant-vowel plus two permissible nasal codas) of Mandarin. The use of many tones serves to offset the asymmetry of a syllable, potentially activating a large number of words. Liu and Samuel cite Maddieson (1978; 2011), which found Mandarin to be among the minority of languages which make use of four or more tones (most only have two or three), including a complex tone (i.e., T3). A chicken-and-egg question then arises: it is not clear whether languages with tone are highly likely to have simpler syllable structures (cf. Maddieson, 2011) or if the tone system has accommodated to the simple syllable structure and potential homophony it (cf. Sereno & Lee, 2015). Jin's (2011) study provides a way out of this issue: in order to avoid the homophony that arises from the simple structure of Mandarin, words in Mandarin tend to be mostly *disyllabic*. Most of the literature has focused on the acquisition of *single syllables and tones*; the aim of the present study is to expand on the questions that can arise in examining the acquisition of *disyllables and tone combinations* e.g., whole-word structures, tone sandhi and neutral tone. A review of the existing acquisition literature follows.

2.2 Phonological acquisition of Mandarin

2.2.1 Segments

In Mandarin, unaspirated stops, fricatives, dentals and retroflexes are said to be the most frequently-used consonants of Chinese syllables (Cheng, 1982). In a review of the phonological development of Mandarin children, Li and To (2017) found a general consensus that plosives and nasals occurred in children's consonant inventories earlier than most fricatives and affricates. Specifically, /p, t, m, h/ were acquired first and the retroflex consonants /ʈʂ, ʈʂʰ, ʂ, ɻ/ were more difficult. In a recent study of the phonological development of affricates in 3- to 5-year-old children, Ma et al. (2022) found a different pattern of development depending on the place of articulation of the affricates: alveopalatal /tɕ, tɕʰ/ were found to be acquired first, followed by retroflex /ʈʂ, ʈʂʰ/, then alveolar /ts, tsʰ/. Li & Munson (2016) examined the development of the voiceless sibilant fricatives in 2- to 5-year-old children and found an acquisition order of /ɕ/, /ʂ/ then /s/.

One of the issues in the acquisition literature has been the definition of what, precisely, constitutes 'acquisition'. As a result, the varying definitions have led different authors to arrive at different ages of child 'mastery' of the components. The general consensus on the acquisition of a component has been 'a certain *percentage* of children producing it *correctly*

at a certain *frequency*'. However, there is no consensus about these italicised keywords. These are discussed in turn.

Researchers have variously used group-level percentages of 70%, 75% or 90% as their criterion for acquisition of consonants. Li and To (2017) adopted group-level criteria of more than 75% or 90% and categorised acquisition/mastery, stabilisation or emergence of sounds for an age cohort, i.e., a sound was considered 'mastered' when more than 75% or 90% of children produced it correctly all the time, 'stabilised' when it was produced correctly two-thirds of the time and 'emergent' when it was produced regardless of target (Amayreh & Dyson, 1998; Sander, 1972; Zhu & Dodd, 2000). Nevertheless, even with regards to this one criterion, 75% and 90% are not one and the same – "more than 75%" is not the same as "90%". Jeng (2013) offered another definition of group-level acquisition, judging a sound as having been acquired if 70% of children in an age cohort produced it correctly in each of the two items in a picture naming task. Some authors have considered individual-level criteria instead, but this has been confounded by the number of opportunities children were given to produce the sounds. Zhu and Dodd (2000) considered a sound as 'stabilised' if it was produced correctly two out of three times whereas Jeng (2011) judged a sound as having been acquired when a child produced it correctly in two tested items. Yet others did not explicitly specify any criteria (e.g., So & Zhou, 2000).

Not only have there been different individual- vs. group-level criteria; different paradigms have also been used to elicit children's responses. Most researchers employed picture naming tasks (e.g., Jeng, 2013; So & Zhou, 2000; Zhu & Dodd, 2000), which meant that pictures and objects served as primes for recall. Mok et al. (2019) note that the result of only administering a small number of words for picture naming means that production performance may be overstated. Meanwhile, Hsu (1996) and Lou (2020) were among the few researchers who examined children's spontaneous speech productions (i.e., no primes offered). Naturalistic studies provide us with a different method of examining children's production patterns because the children are producing the words themselves and are not restricted to certain pictures. Given that the focus of the present thesis is on the segmental and tonal templatic patterns that can be found in the spontaneous speech productions of a small number of children, it also seems most appropriate to adopt the child-specific acquisition criteria that have been developed for identifying templates and which will cater to each of the children's unique developmental patterns. The child-specific acquisition criteria used by researchers in the field of child phonological templates (cf. Vihman, 2019) are thus adopted for this study: a consonant was considered 'acquired' if a child produced it accurately at least twice in different words within a session, and this is considered 'acquired'

from the very first session that this occurs; a consonant was considered 'marginally acquired' if the child produced it accurately only once within a session, and not acquired if the child produced it only as substitutions.

Given the high incidence of fricatives (and affricates) in Mandarin, it also makes sense that this consonantal manner (see earlier discussion in section 1.1) be the segmental phonological contrast of focus in this study. Additionally, when creating stimuli for experiments, e.g., creating nonwords to test children's phonological memory (Edwards et al., 2004; Gathercole, 2006), researchers have also been careful to match the phonotactic probability of occurrence of sounds in a language (cf. Vitevitch & Luce, 2005) in order to emulate the characteristics of the language as much as possible. Thus, manner of articulation is also used as the control factor in the creation of nonword stimuli for an experimental task of nonword repetition in Mandarin in the present thesis (see details in section 3.2.1).

In learning an increasing number of words with increasing complexity, patterns may also be found in consonant sequences of disyllabic words, i.e., how children respond to *two* onsets in tandem. Children's use of word structures (the whole-word characteristics of words) or substitutions in response to challenges in words (also considered 'compensatory strategies' cf. Menn, 1983; O'Seaghda et al., 2013) has been found to differ across languages (Stoel-Gammon & Stemberger, 1994; Vihman, 1978; Vihman, 2019), which indicates effects of the ambient languages. Consonant harmony and reduplication are generally children's preferred structures (Stoel-Gammon & Stemberger, 1994; Vihman, 1978), but they are not universally found across all languages. Vihman's (2019) study of 7 language groups of children at the 25-word point (US English, UK English, Estonian, Finnish, French, Italian and Welsh) found that consonant harmony was prevalent amongst all groups except Welsh (also see one of the earliest studies, Vihman (1978), which showed low rates of consonant harmony in Cantonese compared to other languages. Kehoe (2015) reported on the word templates used in three Spanish- or German-learning children at the 25wp: there was a higher proportion of variegation by children learning Spanish than those learning German.

2.2.1.1 Identification of templates in consonant sequences and word structures

The most recent systematic study of acquisition of segments and word structures in children learning Mandarin was Lou (2020). Lou looked at Mandarin word production in children between the beginning (the '4-word point', 4wp) and end (the '25wp') of the single-word period and tried to examine if there was evidence of templatic usage in the children's words.

The 4wp is the first month in which the child spontaneously and reliably produces four or more adult-based words in a half-hour session, and has been established in the literature as the beginning of lexical use (Vihman, 1996). The 25wp then marks the end of the single-word period: it is the first month in which the child spontaneously and reliably produces 25 or more words and may also begin to combine words. It is an important milestone as this is where phonological templates may begin to be observed in children's speech (Vihman, 2019, as mentioned in Chapter 1).

In terms of segments, at the 4wp (roughly 13 months of age, in Lou's study), most of the children produced /p, t/, a few other children additionally produced /m, n/ and one child additionally produced /ɕ, tɕ/. At the 25wp (roughly 18 months of age), all the children produced these consonants more reliably, with the proportion of fricatives/affricates being higher than that of nasals, reflecting the proportions in the language. This pattern of findings resembles those of Li and To (2017). Certain children also either selected /k/-, /t/- and /tɕ/-initial words or adapted other words to fit these onsets: this was presented as evidence for templatic use at the 25wp. In terms of word structures, the children made the most use of CVglideV or CVC. Child forms at both the 4wp and 25wp were mostly either reduplicated or variegated, reflecting the frequency of these structures in the mothers' input. Consonant harmony was rare; children learning Mandarin do not tend to make use of harmony for variegated targets (see discussion in Vihman et al., 2022). The templatic patterns that might occur in consonant sequences at the end of the single-word period/beginning of the word combination period have yet to be established. These will be explored in this study.

2.2.2 Tone

In general, tone has received more attention than segments in the developmental literature on Mandarin. The acquisition of tone has long been of interest to researchers (e.g., Clumeck, 1980; see Singh & Fu, 2016 for detailed review). In the perception literature, researchers have studied the discrimination of lexical tones by tone learners (e.g., Tsao, 2017; Yeung et al., 2013) vs. non-tone learners (Götz et al., 2018; Liu & Kager, 2014; Mattock & Burnham, 2006) and the lexical integration of tone by tone learners (Gao et al., 2011; Singh et al., 2015) and non-tone learners (Quam & Swingley, 2010; Singh et al., 2008). Most studies on Mandarin production have been geared towards establishing the order of tone acquisition. Some claim that T1 and T4 are acquired before T2 and T3 (Li & Thompson, 1977; Lou, 2020, Zhu, 2002); Wong (2012, 2013) proposed a correlation between order of acquisition and articulatory complexity: T4 > T1 > T2 > T3, while others

provided yet a different order of tone acquisition along with ages of acquisition: T1 = T2 (2;0) > T4 (2;6) > T3 (3;0) (So & Zhou, 2000).

Of more interest and relevance to the present study are *production* studies examining when children can be said to have attained 'tone mastery' (see further discussion of 'mastery' below) (e.g., Clumeck, 1977; Li & Thompson, 1977; Wong & Strange, 2017; Xu Rattanasone et al., 2018; Zhu, 2002). The first recorded Mandarin production study was Chao's (1951; cited in Wong et al., 2005) study of his 28-month-old granddaughter's Mandarin language abilities. Chao did not provide details about her tone production except that tones were produced correctly and that she had some difficulties with the tone sandhi rules.

As with the order of acquisition of consonants, the purported age of 'mastery' has also been inconclusive, as different groups of researchers administering different paradigms have found age differences. Using picture naming tasks, some researchers found that children aged 1.5 to 4 years make few errors on tone (Zhu & Dodd, 2000) or have mastered tone acquisition by the time they are aged 2 to 3 years (Li & Thompson, 1977). Zhu's (2002) longitudinal study of children from 10 to 24 months revealed accurate production of all four tones before age 2, but Clumeck's (1977) longitudinal study of three children between 1 and 3 years of age found that tone production had not been mastered at age 3 years. Once again, it is unclear what 'mastery' meant (e.g., a certain percentage match to target, zero error etc.). Furthermore, Li and To (2017) stated that many of the authors of the studies they reviewed did not discuss tone acquisition at all as their participants were older than 2.5 years and tone was assumed to have already been mastered by then.

While there are at least some criteria for the acquisition of segments in Mandarin (see section 2.2.1 above), no one seems to have proposed a corresponding percentage of occurrence for Mandarin tones (although Mok et al., 2019 report that So and Dodd, 1995, defined a tone as being acquired in Cantonese when it was "used contrastively on at least 50% of opportunities or correctly on 90% of opportunities" [p. 484], without further explanation). The general trend for studying tone seems to be to measure how *adult-like* children's productions of tones are. Using a combination of experimental methods – Wong and colleagues (Wong et al., 2005; Wong, 2012; Wong & Strange, 2017) elicited tone productions by children aged between 2 and 6 years and had adults make judgements of these productions; Xu Rattanasone et al. (2018) administered a picture naming task to children aged between 3 and 5 years then ran acoustic analyses on their productions – these authors all report inconsistent and non-adult-like tone productions even in children up to 6 years of age. A similar pattern of inconsistencies has also been reported for Cantonese

– earlier studies (e.g., So & Dodd, 1995; To et al., 2013) claim that children produce tone accurately by age 2;0 or 2;6 but more recent studies by Mok and colleagues (2019) have also found non-adult-like tone productions even in 6-year-olds.

If 'mastery of tone' is thus defined as how similar the tones children produce are to the target, then the physiological limitations in speech motor control may be a reason why tones are not adult-like even by age 6 (Munson et al., 2011; Li & Munson, 2016; Wong, 2012; Wong et al., 2017). Tone is a function of the rapid modulation of the vocal cord muscles, which tense or relax, depending on the tone (Ohala, 1978; Sagart et al., 1986; Wong, 2012) and precise speech motor control is needed to differentially produce acoustically similar tones (Wong et al., 2017). It is unlikely that children are fully equipped yet with the hardware needed to adjust the differences among tone categories by age 6 as the vocal folds are not fully developed until adolescence (Crelin, 1987; Kent & Vorperian, 1995; cf. Li & Munson, 2016 oromotor maturation hypothesis). Some tones are easier to produce than others, for example, T4 is articulated by decreasing vocal fold vibration or releasing the tension of laryngeal muscles, resulting in a natural falling contour. The process of acquiring tone is said to be a protracted one, and influenced by the production intricacies of the respective tones (Wong, 2012); Wong (2013) thus proposed the following order of acquisition, as mentioned earlier, which corresponds to increasing complexity: T4 > T1 > T2 > T3.

However, Mok et al. (2020) contend that if successful acquisition of tones is to be based on comparing children's emerging productions against adults, who presumably have "ceiling accuracy" (p. 12), then that might underestimate children's abilities. They propose that successful acquisition be considered as children's meaningful use of tones instead, i.e., above chance level; this allows room for development as a result of production practice in adult-like accuracy (cf. maturing speech motor components, Li & Munson, 2016) and fine-tuning among the tone contrasts. This above-chance level definition will be adapted for the present study to suit disyllabic Mandarin stimuli (see details in section 3.3.3.2).

Most researchers have focused on the acquisition of *single tones*; fewer have examined the developmental trajectory of *disyllabic tone sequences*. Studying children's productions of disyllables allows for the study of words that may undergo T3 sandhi processes and the acquisition of T0 (which typically does not occur in isolation, Zhu, 2002). In perception research, children have been found to reliably *comprehend* T3 sandhi forms from age 3 years (e.g., Chen et al., 2010; Wewalaarachchi & Singh, 2016), while researchers studying production have found that children of a similar age do not yet reliably *produce* them (Wang, 2011).

Developmental differences have also been observed in a productive application of the T3 sandhi process to novel disyllabic or trisyllabic word structures: Tang et al. (2019b) found that 3-year-olds could apply the rule to disyllabic word structures, but that 5-year-olds still performed differently from adults when they encounter trisyllabic word structures that could be left- [(σσ) σ] or right-branching [σ (σσ)]. However, even adults were found to be less consistent in applying T3 sandhi rules than the half-T3 sandhi rule (i.e., just the falling 21 pitch of the 214 contour) that typically occurs with words preceding T1, T2 or T4, in response to disyllabic 'wug words' made up of accidental gaps in the Mandarin syllabary (i.e., phonotactically possible syllables that happen not to be existent words, cf. '*tonotactic* accidental gaps' in section 3.2.1, on the creation of nonword repetition stimuli) (Zhang & Lai, 2010). In general, both children and adults make use of the heuristics that the language offers – lexical knowledge, morphological structure and frequency of occurrence of sandhi forms – to guide their productive sandhi responses (Tang et al., 2019b).

Furthermore, T0 is said to be acquired later than the lexical tones, and to be potentially mastered only by 4;6 (Tang et al., 2019a). Chen and Xu (2006) posit that given the characteristics of T0, it is not surprising that it should be challenging for children to acquire. Based on their thorough investigation of F0 contours of T0 syllables, Chen and Xu postulate that T0 has a mid-level pitch (about the midpoint of the highest and lowest F0 values of the full tones) independent of its neighbouring tones, but its weak articulatory production is probably the cause of the large amount of variability observed. Their regression analyses reveal that the variations found in the F0 of the T0 syllable were mostly due to the F0 height and the velocity of and from the *preceding* full tone (cf. Xu, 1997). T0, being produced with a weaker force than the other tones, is subject to more influence from the preceding tone. Furthermore, T0 syllables are only half the length of full tone syllables (Lin & Yan, 1980, cited in Chen & Xu, 2006), so a lack of time limits the child's ability to reach the target.

In the acquisition literature, Zhu and Dodd (2000) report that no three-year-olds successfully produced all T0 words accurately, with some even omitting the T0 syllable entirely; less than 40% of 4-year-olds produced T0, indicating a longer and slower rate of development for T0. Errors with the T0 syllable include the substitution of T0 by a full lexical tone (a phenomenon also observed in Singapore Mandarin: cf. Shang & Zhao, 2013, 2017; see section 2.4 below) or the lengthening of the usually-shorter T0 syllable (Li & Thompson, 1977; Tang et al., 2019a; Zhu & Dodd, 2000). Once again, developmental differences were observed in a productive application of T0 to novel word forms: Tang et al. (2019a) required children to generate new disyllabic words with a lexical tone and neutral tone. They found that 3- and 4-year-olds could apply the neutral tone category in these novel word formation tasks, but the

productions were not yet adult-like (cf. Wong et al., 2005; Wong, 2008). The younger children produced a more falling pitch contour for the neutral tones, seemingly producing them as the falling T4 instead.

2.2.2.1 Identification of templates in tone sequences

A recent approach to studying tone sequences has been Mok and Lee's (2018) adoption of the idea of 'phonological templates' (Vihman & Croft, 2007; Vihman, 2019) for identifying tone templates in Cantonese. Mok and Lee (2018) first identified this phenomenon in their study of five 2- and 2.5-year-old children simultaneously learning Cantonese and English in Hong Kong. They analysed the tone productions of the children's speech in both Cantonese and English and found a difference between the three Cantonese-dominant bilingual children and two who were less proficient in Cantonese. The latter group seemingly adopted a 'high-low' template (T1-4) for predominantly reduplicated disyllables in Cantonese. The authors proposed that this was due to cross-linguistic transfer of the natural drop contour of the stress patterns in trochaic English words (e.g., **mo**ther, **bro**ther, stressed syllable in **bold**), which make up 90% of English disyllabic words (Cutler & Carter, 1987) to the children's Cantonese productions. In a further analysis, the authors found that the more Cantonese-dominant children also produced the 'high-low' template at an earlier point, when their Cantonese was less proficient.

Mok and Lee (2018) acknowledge that the 'high-low' template found in their study was similar but not identical to the phonological template proposed by Vihman and colleagues. Mok and Lee explain several differences between the two:

- Vihman's phonological template referred to "certain *segmental* combinations" (p. 15, italics added) while the 'high-low' template was a tone template and therefore a *suprasegmental* feature. Given that the initial samples on which the term 'phonological template' was introduced was first based on non-tone languages, the lack of reference to suprasegmental features like tone is understandable. Nonetheless, "prosodic and rhythmic structures of the adult languages can also influence the shapes of the templates" (p. 15), which gives the option of expanding the term 'templates' to cover 'tone templates' (and other features) as well.
- Vihman's phonological template has been found mainly in children who were at the one-word stage whereas Mok and Lee's children were beyond that stage when they produced the template. However, the phonological templates that have been identified have from previous work have been on monolingual populations or bilingual

populations learning two non-tone languages. Perhaps learning English together with a tone language like Cantonese, two languages with such contrasting phonological structures, extends the length of time on which children may still be using templates to support their word production.

Nonetheless, Mok and Lee propose that the co-occurrence of both 'inaccurate' (i.e., use of templates) and accurate (adult) forms fits in with the idea that phonological development is a U-shaped process where, rather than representing the loss of an ability or system, phonological abilities are being *reorganised* with language use and learning (cf. Werker et al., 2004). This is in line with Vihman's phonological template definition that templates are constantly evolving and reorganising depending on the child's phonological repertoire (see Chapter 1).

Mok and Lee (2018) conclude that the 'high-low' template in their study could be viewed as a subtype of Vihman's phonological template. The authors also highlight that their template occurred predominantly in reduplicated words. In the Chinese languages, reduplication is a morphological process (cf. Li & Thompson, 1981) often used in child-directed speech (Matthews & Yip, 2011; but see the discussion sections of 4.1 and 4.2 for reduplication in Mandarin that extends beyond child-directed speech; cf. also Lin, 1990). Mok and Lee suggest that the children may be applying a diminutive morphological template to a tonal pattern that is similar to diminutive forms in English (e.g., *kitty*, *doggy*), always produced in a trochaic ('high-low') pattern. The authors urged future researchers to explore if the 'high-low' template (or potentially other tone patterns) could be found beyond reduplicated forms.

In a pilot of the present study, data from two children in Yorkshire, UK, was analysed (Choo et al., 2019). These children were exposed mainly to Mandarin at home and to English in the wider society. There was a prevalence of a T1-0 tone template among the children's Mandarin words (Choo et al., 2019), similar to but different from Mok and Lee's high-low T1-4 template, but the template was not exclusive to reduplicated forms, unlike Mok and Lee's predominant finding. Children in Singapore are exposed to multiple languages at home and in the environment, including English and other Chinese languages (see Chapter 3 for further discussion). Hence it would not be surprising to see a cross-linguistic element of English stress patterns being adapted to Mandarin words in the present study as was found in Mok and Lee (2018) and in the pilot (Choo et al., 2019).

As mentioned earlier (section 2.2.1.1), the most recent systematic study of children's word productions in Mandarin between the 4wp and 25wp by Lou (2020) also examined the acquisition of tone and tone templates. In terms of tones, Lou found monosyllabic tone templates in her children's productions at the 4wp but not the 25wp. At the 4wp, T1 and T4 were selected for production as well as substituted for the other tones, leading to high frequency of usage with moderate precision. The high perceptual salience of these tones and the considerable babbling practice children have had with them (cf. Lou et al., 2018) led to their using and over-using the tones. However, these were no longer templatic by the 25wp. For disyllabic sequences, Lou found that children favoured two tone sequences with T1 in the first syllable (T1-4 and T1-0, the first of which is similar to the Cantonese T1-4 pattern found in Mok & Lee, 2018), but Lou deemed these two tone sequences as not being templatic at either the 4wp or 25wp time points because they were used mostly as substitutions. It remains to be seen whether disyllabic tone templates can be found at the end of the single-word period/beginning of the word combination period in Mandarin.

2.3 The (a)symmetry of segmental and tonal information in Mandarin

Work with a diverse group of languages (e.g., Chinese languages: Mandarin, Cantonese; Germanic: English, Dutch; Japanese; Romance: French, Spanish) has shown that the phonological structures of languages shape how words are represented in the respective languages. Speakers of different language groups use different units to segment speech into words, e.g., stress influences the syllabification in European languages (e.g., Cutler et al., 1986 on English and French; Sebastian-Galles et al., 1992 on Catalan and Spanish) and the preferred units of processing are syllables in Mandarin (e.g., Tseng, 1998; Feng et al., 2019) and morae in Japanese (e.g., Otake et al., 1993). These findings are corroborated by evidence from speech error analyses which show that errors reflect the characteristics of the language (some errors associated with consonants and vowels can be used to explain tone errors in Thai, Gandour, 1977; whole-syllable errors are rare in English but not in Mandarin, Chen, 1993; 2000; morae rather than segmental errors are found for Japanese, Kubozono, 1989), and in experimental studies which use novel words that adhere to the phonotactics of a language (e.g., Dell et al., 2000; Smalle et al., 2017).

Within the Mandarin syllable, researchers have found an asymmetry in the processing of segmental and tonal information. Work with adults testing the dissociation of syllable vs. tonal processing in perception studies have made use of priming paradigms that differ in a range of factors (e.g., task, inter-stimulus interval, position of overlap between prime and target and lexical status of primes, cf. Zwitserlood, 1996). Some authors found that

segmental information was responded to more quickly than tonal information (e.g., Chen et al., 2002; Lee, 2007). A relative tonal advantage was then found in situations where top-down processing was involved (e.g., in the presence of white noise or having to make judgements of words when embedded in sentential or idiomatic contexts, cf. Liu & Samuel, 2007; Sereno & Lee, 2015). The few production studies (e.g., Qu et al., 2012; Yu et al., 2014) show an advantage of segments as opposed to the predominantly judgement-based perception tasks that show a syllable over segment advantage. No study of production has compared segmental information and tonal information so that remains to be done.

How does one reconcile these opposing findings? Mandarin has a simple syllable structure, so any change in the individual segments may cause a disproportionate amount of change, unlike, e.g., English. Thus, it may be difficult to compare the findings from different studies all modifying different parts of the stimuli to different extents. This is especially so for the studies manipulating the onsets of words, which have been shown to be a crucial unit of processing, especially in models of word recognition in European languages that postulate a serial order of processing (cf. Shattuck-Hufnagel, 1987, Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986) and some researchers on Mandarin who have also reported successful primes with word onsets (e.g., Qu et al., 2012; Yu et al., 2014). Studies of English phonological errors show that the resultant item is often a real word (e.g., Fromkin, 1973); this finding is likely more pronounced in Mandarin, where any change to one or more segments is all the more likely to result in another real word as well (Wan & Jaeger, 1998). Indeed, according to Deng and Dang (2007; see also Zhu, 2002 who cited the *Xiandai Hanyu Cidian* 'Modern Chinese Dictionary'), disregarding tone, there are only about 400 unique syllables in Mandarin while there are about 9,000 in English (Huff, 2017; Levelt et al., 1999 cite 12,000 unique syllables in English).

In the acquisition literature for Mandarin, a group of researchers have found developmental differences in the asymmetry of segmental and tonal information. In a series of mispronunciation studies using the preferential looking paradigm, Wewalaarachchi and colleagues (Singh et al., 2015; Wewalaarachchi et al., 2017; Wewalaarachchi & Singh, 2020) have found developmental differences (toddlers vs. preschoolers vs. kindergarteners) and population differences (monolingual Mandarin-speaking vs. bilingual Mandarin-English-speaking) in sensitivity to segmental (consonant and vowel) and tonal mispronunciations. In all of these studies, the naming effects in four conditions (correct pronunciation, consonant mispronunciation, vowel mispronunciation and tone mispronunciation) were calculated. Singh et al.'s (2015) study was on bilingual-Mandarin-English-speaking children; they found that toddlers (2.5- to 3.5-year-olds) were more sensitive to tone than to consonant and vowel

mispronunciations, i.e., they found it more difficult to recognise the target when tones were mispronounced than when consonants or vowels were mispronounced. In contrast, preschoolers (4- to 5-year-olds) were more sensitive to consonant and vowel mispronunciations than to tonal ones, seemingly becoming less sensitive to tone information as they got older. Perhaps the acoustically salient tone helps younger children learn words early on, but as their lexicon grows, segmental information becomes more important.

Two follow-up studies examined the age and population effect in more detail. Wewalaarachchi et al. (2017) tested monolingual Mandarin-speaking and bilingual Mandarin-English-speaking 2-year-old children and found that the monolinguals demonstrated equal sensitivity to vowel and tone mispronunciations and least sensitivity to consonant mispronunciations. On the other hand, the bilinguals demonstrated a decreasing order of sensitivity to vowels, consonants, then tones. Wewalaarachchi and Singh (2020) examined the performance of monolingual Mandarin-speaking and bilingual Mandarin-English-speaking 6-year-old children and once again showed that both groups were more sensitive to consonant and vowel mispronunciations than tone mispronunciations. They concluded that tone sensitivity seemingly attenuates as children get older, independent of whether they are from a monolingual or bilingual language background. Implementing a different experimental paradigm of novel word learning with monolingual Mandarin-speaking 2- and 3-year-olds, Ma et al.'s (2017) findings corroborate those of Wewalaarachchi and colleagues': Ma et al. also found a decline in tone sensitivity with age. At an older age, vowel information becomes more important than tone information and remains the most constraining source of lexical contrast (cf. Tong et al., 2008).

The kinds of dissociation between segmental and tonal information in Mandarin-learning children in Singapore have not yet been systematically studied. Children in Singapore are exposed to multiple languages at home and in the environment. The bilingual education policy dictates that children grow up bilingual and learn to speak English and an official 'mother tongue' of Mandarin, Malay or Tamil (determined by father's ethnicity, regardless of actual usage at home, cf. Stroud & Wee, 2010). The children may also be exposed to other Chinese languages spoken by the grandparent or great-grandparent generations, who speak only those languages. A discussion of Mandarin as it is used in Singapore follows.

2.4 Mandarin in Singapore

Mandarin in Singapore is known as 'Huayu'. The Singapore government deliberately adopted a different name, a departure from the 'Putonghua' used in Mainland China so as to

afford some distance between the two countries (Guo, 2007; Shang & Zhao, 2013). Nonetheless, the variety of Mandarin taught in Singapore emulates the Beijing Mandarin variety of Putonghua in China. Again, due to reasons relating to political sensitivity, the Singapore government has never explicitly mentioned using Putonghua or Huayu as the standard in Chinese language education and assessment (Shang & Zhao, 2017). Again, nonetheless, when Mainland China changed their writing system to the simplified script and introduced the use of *Hanyu Pinyin* to aid their teaching instruction of the pronunciation of Chinese characters in 1973 (Ang, 1994; Shang & Zhao, 2013), Singapore followed suit. Interestingly, there are more similarities between the phonology of Singapore Huayu and Taiwan Guoyu, due to the fact that a large proportion of the Chinese population in both countries speak Hokkien (one of the endogenous languages in Singapore, Lim, 2007; or 'Taiwanese', as termed in Taiwan; Hokkien and Taiwanese are both dialects of Southern Min) and the popularity of Taiwanese drama series, pop music and entertainment programmes in Singapore (Chong & Tan, 2013; Shang & Zhao, 2017). Because the use of 'Huayu' is only familiar to Singaporeans, the general term 'Singapore Mandarin' will be used throughout the rest of the thesis except when comparing with other varieties.

Shang and Zhao (2013, 2017) proposed that Singapore Mandarin can be categorised along the acrolect-mesolect-basilect continuum (cf. Platt & Weber, 1980). The acrolectal variety of Singapore Mandarin resembles Beijing Putonghua in most linguistic aspects, with the exception of uniquely Singaporean lexis. It is the language used in education, the media and proficient speakers in formal contexts. The mesolectal variety is the one predominantly used by Singaporean Mandarin speakers on a daily basis in non-formal engagements. The basilectal variety diverges the most from Putonghua, with its influence from the other languages in the environment and large amounts of code-switching among these languages.

2.4.1 Segments

The "core" (cf. Chew, 2007) of the phonology of Singapore Mandarin is Beijing Putonghua with additional uniquely Singaporean attributes (Chen, P., 1999; Shang & Zhao, 2013). Based on the video recordings from Singaporean children and their parents, Chen (1986) delineated the salient segmental features in Singapore Mandarin and found different realisations of the pronunciations of Mandarin in Singapore and Beijing Putonghua. The differences in varieties are presented in Table 2.1.

Table 2.1: Differences in realisations of consonants and vowels in Beijing Putonghua and Singapore Mandarin (from Chen, 1986)

Beijing Putonghua	Possible realisations in Singapore Mandarin
tʂ	tʂ, ts
tʂʰ	tʂʰ, tsʰ
ʂ	ʂ, s, ʃ as in English
ɻ	ɻ, dz*[3], l*
ɕ	ɕ, s
x	h as in English, f*, w*
l	l, ɻ, n*
initial n	n, l*
final n	n, ŋ
final ŋ	ŋ, n
ɚ[†]	ə
ɤ	ə
ɛ	e
iɛ	iɛ, ie
uo	uo, uɔ
yɛ	yɛ, ye
iou	iou, iu
uei	uei, ui
iɛn	iɛn, ien
yan	yan (very rare), yɛn, yen
uŋ	uŋ, oŋ, ɔŋ
iuŋ	iuŋ, ioŋ, iɔŋ

* see footnote; † rhotacisation; see in-text below

However, as Chen (1986) was approaching the issue from a pedagogical standpoint and comparing the Singapore Huayu pronunciation against the Putonghua standard, the latter of which the large majority of the Singapore population do not speak, she deemed the deviations 'errors'. For instance, Chen (1986) found that Singaporeans did not use retroflexed consonants or vowels very reliably, as did Lock (1989), but this is not surprising as the use of retroflexion is rare even in Mainland China outside of Beijing, the variety on which Putonghua is based (Chen, 1986; Lock, 1989; Shang & Zhao, 2017; Zhu, 2002). Lock (1989) notes that many of the features that do not appear in Singapore Huayu or Taiwanese Guoyu do not occur in Mainland China outside Beijing either. Even though the Mandarin television broadcasting company (then *Radio Television Singapore*, now *Mediacorp*) recruited Mainland Chinese as television and radio newscasters (Chen, 1986; Shang &

---

[3] Anecdotally, all cases marked with * are more prevalent in people in Singapore aged over 60. For instance, in this study, Fu Zhen's grandmother, who was her interlocutor, produced *ɻən2* 'person' as *lən2*. The ɻ to l substitution is likely cross-linguistic influence from the corresponding Hokkien term *laŋ2*.

Zhao, 2013) and teachers (Lock, 1989) and Singapore has since seen an increasing number of Mainland Chinese immigrants in the service and labour industries, the predominant variety of Mandarin spoken in Singapore is still Huayu, not Putonghua. Additionally, as previously mentioned, there is more resemblance between Singapore Huayu and the Taiwanese Guoyu than Beijing Putonghua.

The use of rhotacisation in Beijing Putonghua is also said to be infrequently occurring in Singapore Mandarin (Lock; 1989; Shang & Zhao, 2013). Rhotacisation is unique to Beijing Putonghua, may appear in words that are "common, familiar or small" (Zhu, 2002, p. 38), and may carry diminutive or negative connotations.

2.4.2 Tones

The use of neutral tone (T0) in Beijing Putonghua is also said to be infrequently occurring in Singapore Mandarin (Lock, 1989; Shang & Zhao, 2013). The previous finding for the lack of use of T0 in Singapore Mandarin is similar to Chen and Kent's (2009) finding in Taiwanese Mandarin. For Taiwanese Mandarin, T0 are commonly produced as short mid-falling tones. For Singapore Mandarin, syllables that are typically realised with the T0 are said to be produced as one of the four basic tones (Shang & Zhao, 2013). For example, based on the regular and irregular weak stress categories Zhu (2002) delineated (see section 2.1.2), Singapore Mandarin speakers would produce particles (e.g., *la4*, although different tones attributed to 'la' affords different pragmatic meanings, see Lim, 2007), affixes (*pi2tsʅ0* 'nose' → *pi2tsʅ3*), reduplicated forms (*ɕiŋ1ɕiŋ0* 'star' → *ɕiŋ1ɕiŋ1*), and lexeme-dependent forms (*tʰou2fa0* 'hair' → *tʰou2fa3* or *tʰou2fa4*) non-neutrally. Indeed, the mothers and children in the longitudinal study produced various items non-neutrally, for example, for the *tɕiɛ3tɕiɛ0* 'older sister' example (provided under the discussion on 'Tone sandhi' in section 2.1.2, although it could also apply to the discussion on neutral tone as it is relevant here), mothers and children produced it in myriad different ways, including: *tɕie3tɕie1*, *tɕie3tɕie2* and *tsɛ3tsɛ2* (both mothers and children), *tsɛ3tsɛ1* and *tsɛ3tsɛ2* (mothers only), and *tɛ1tɛ1*, *tɕe1tɕe1* and *te3te2* (children only) [examples taken from Tables 4.14 and 4.15]. Other examples are discussed further in section 4.2.

Scholars have also noted differences in the tone realisations of Singapore Mandarin and other varieties. Chua (2003, cited in Lee, 2010) notes that the tones in Singapore Mandarin may appear 'lower' in general, presented in Table 2.2.

Table 2.2: Tone differences between Putonghua and Huayu using Chao (1931) numerals (Chua, 2003; cited in Lee, 2010)

Tone	Beijing Putonghua	Singapore Huayu
1	55	44
2	35	24
3	214	211
4	51	41

However, Lee (2010) states that Chua (2003) did not specify how the numbers were assigned and that these were likely to be impressionistic judgements. Lee (2010) ran an acoustic study to examine the realisations of the four tones. They found that tone influenced syllable duration, in that T4 was significantly shorter than the other three tones, which were not significantly different from one another. T4, the shortest tone in Singapore Mandarin, was said to be more similar to Beijing Putonghua than Taiwan Guoyu, however, Tones 1, 2 and 3, which are similar in duration, were different from Beijing Putonghua (see below).

Lee (2010) depicted the tonal contours of the four tones in Singapore Mandarin in Figure 4 of their study, reproduced as Figure 2.1 below:
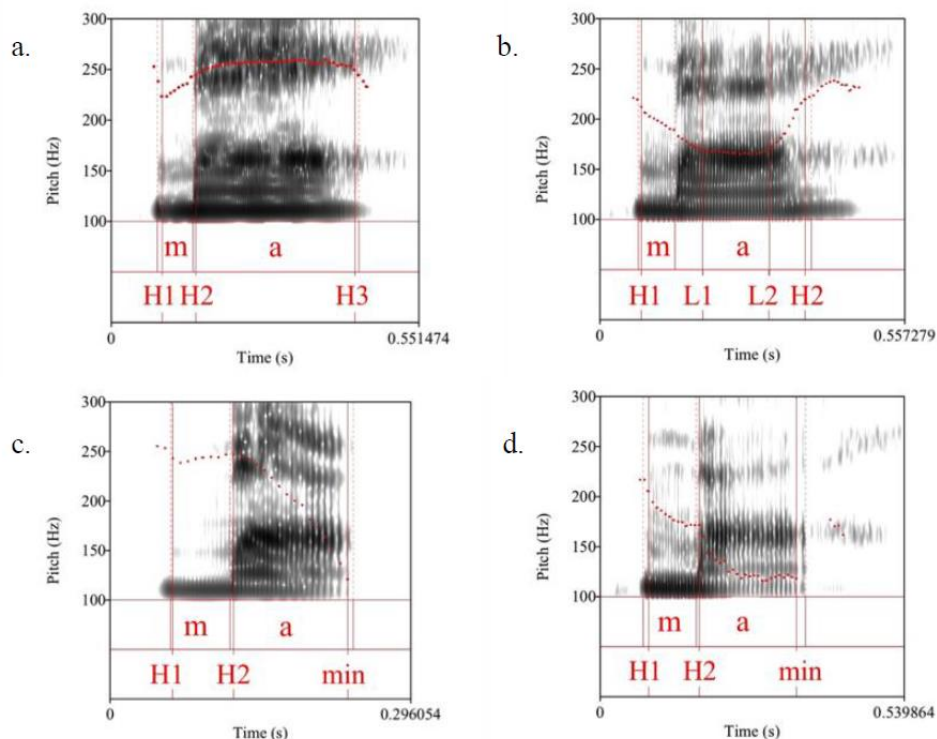
Figure 2.1: Tonal contours of the four tones in Singapore Mandarin: (a) T1, (b) T2, (c) T4 and (d) T3, reproduced from Figure 4 of Lee (2010)

Based on Lee's (2010) analysis, T1 and T4 start at a high pitch, with T1 remaining high throughout the syllable while T4 is high through the onset but then falls to a low level. T2 and T3 begin at a mid-high level, with T3 falling to a low level. T3 in Singapore Mandarin is more similar to T3 in Taiwanese Guoyu than in Beijing Putonghua: T3 does not end low, it is realised as a dipping tone instead. The differences between T2 and T3 in Singapore Mandarin have previously been reported to be negligible (Xu & Wang, 2004), which is not surprising, given that Wang and Li (1967) showed in their tone perception experiment that even adults could not tell these two tones apart. The interesting finding in Lee's (2010) study was that, with T2, there was a fall to a mid-low stretch before a rise to a high pitch level at the end of the syllable – this mid-low stretch has not been recorded in other varieties: T2 in Beijing Putonghua falls slightly before a rise (Xu, 1997), while T2 in Taiwanese Guoyu remains level or dips slightly before rising towards the end (Shih, 1988, as cited in Lee, 2010). Acoustic contours are further presented and discussed in Chapter 4.

Given that past segmental and tonal comparisons of Singapore Mandarin with other varieties have been with adults, the segmental and tonal realisations of children learning Singapore Mandarin from adults who themselves may be speaking with these 'uniquely' Singapore Mandarin realisations has not yet been recorded. The present study is an attempt at examining the segmental and tonal representations of children learning Singapore Mandarin.

To recapitulate, the aim of the present study is to examine the possible use of segmental and tonal patterns in the disyllabic words of children learning Singapore Mandarin, based on the whole-word phonology approach. Children's speech responses are studied in two ways: by means of naturalistic observations of everyday spontaneous and imitated speech and an experimental nonword repetition task in Mandarin. The study of the influence of long(er)-term representations in children's knowledge is thus complemented by a study of short-term retrieval and production practice in nonword repetition. A discussion of the methodology follows in Chapter 3.

Chapter 3 – Methodology

3.1 Participants

3.1.1 Naturalistic study

Six children were video- and audio-recorded monthly in Singapore. Two children used mainly English in the sessions and so were excluded from subsequent analyses. The children and their parents were born in Singapore. Typical development was established through parent report. The intention was to record them at the end of the single-word period and the beginning of the word combination period, usually identified as the first session in which children spontaneously produce 25 or more uniquely identifiable word types in a single half-hour recording session (the 25-word point, or 25wp, Vihman & Miller, 1988, see Chapter 1 for discussion). Five recording sessions followed that first 25wp session, so most of the children were recorded for six sessions[4]. Families were paid for their time after each session.

3.1.2 Experimental study

Twenty-three 2-year-old children were recruited via word-of-mouth and social network platforms in Singapore. Three children were deemed unsuitable (one parent did not interact with their child; the other two children did not produce any responses for the nonword repetition task), leaving the final sample of 20 children. The children and their parents were born in Singapore. Typical development was established through parent report. Families were paid for their time after each session.

3.2 Materials

Information sheet and consent form: The ethics committee from the Department of Language and Linguistic Science, University of York approved the study before data collection commenced. Parents were given an information sheet detailing the study (in English and Chinese) and consent forms (in English and Chinese) to sign.

Vocabulary checklist (Singapore Communicative Development Inventory [CDI]; Singh et al., 2022): After each recording the mother was given the CDI form fill in. This was a bilingual form that had English and Chinese words on it. The mothers checked off the words their

---

[4] One child, En Ting, had only five sessions in total due to scheduling conflicts.

children understood and/or produced in Mandarin and English. The CDI served as a measure of words children understand *and* say, so it is a good proxy for input to the children as well as for child output.

The CDI was also used as a measure of ambient frequency of the language that the children are exposed to: 1) the words were classified into categories of word structures (see definitions in Table 4.4) for a measure of the ambient frequency of word structures in child-directed speech; 2) the ambient frequency of tone combinations in child-directed speech was also calculated. Reference to these frequency counts in the CDI will be made where relevant.

Language background questionnaire (LBQ, Lee & Rickard Liow, 2015): The primary caregiver (the mother in all cases) was given an LBQ to provide information on the languages the child used to communicate with all caregivers, the age the child was first exposed to the languages and the child's proficiency in each language. Families were also asked to provide information on how many hours each caregiver spent with the child and how many hours the child spent in a nursery/child care centre, if any. Most childcare centres in Singapore are standardised – the centres conducted their classes in Mandarin and English for an equal number of hours each.

For the naturalistic study, of the four parent-child dyads, two parents spoke Mandarin to their children exclusively, the other two spoke both Mandarin and English. Three families had a helper at home who came from the Philippines or Indonesia and spoke English to the children. For the experimental study, of the 20 families, 8 parents spoke Mandarin to their children exclusively, the remaining 12 spoke both Mandarin and English. Fourteen had a helper at home who came from the Philippines or Indonesia and spoke English to the children. Regardless of Mandarin-dominant or balanced bilingual input, there were no discernible differences between the two groups – all the children produced the patterns found in the study, with some potential idiosyncratic preferences that related to the child rather than the input (e.g., the vowel-consonant-vowel [VCV] syllable structure of Yan Min's real name might have primed her to produce more VCV words).

All the children either had grandparents living with them or visited their grandparents on the weekends. The grandparents spoke in Mandarin or one of the other Chinese languages (Hokkien, Cantonese and Teochew) to one another. Some grandparents spoke exclusively in Mandarin to their children and grandchildren; others spoke in both Mandarin and English. The ambient language inside and outside of home was thus a mixture of English, Mandarin

and other Chinese languages. Many children are thus simultaneous bilinguals, with proficiency in English and Mandarin (and possibly other Chinese languages, depending on whether the people who speak these languages serve as caregivers) differing by family.

3.2.1 Experimental study: Nonword repetition stimuli

In section 1.3.2, several issues with extant nonword repetition tasks were discussed. In order to create nonwords that better match the frequency of the phonological structures of the language, the frequency of segmental sequences and tone sequences of existing disyllabic words were first calculated using the distributional frequency of the disyllabic words in the Singapore Communicative Development Inventory (CDI; Singh et al., 2022).

The CDI served as a measure of the words children are exposed to in Singapore. Disyllabic words were used because most words in Mandarin are disyllabic (70%; Institute of Language Teaching and Research, 1986, cited in Zhou & Marslen-Wilson, 1995; Jin, 2011; Zhu, 2002). This accords with the CDI: of the 634 words in the Singapore CDI, 372 (59%) were disyllabic [195 (31%) were monosyllabic, 65 (10%) were longer than 2 syllables and 2 (<1%) were English-only with no Mandarin equivalents].

The distribution of segmental sequences in the CDI was examined first in order to establish the segmental sequences that exist in the language and used to guide the creation of the nonwords in the nonword repetition task. Segmental sequences refer to the manners of articulation of the onset consonants of the two syllables of the disyllabic nonwords. In Mandarin, 11 of 21 (52%) consonants that can occur in the onset position are fricatives (and affricates, classed together as 'fricatives', F), followed by 6 plosives (P, 28%), 2 nasals (N, 10%) and 2 liquids (L, 10%). This manner of articulation proportion revealed that fricatives dominated onsets, and indeed, the frequency of occurrence in the CDI follows this: 172 of 372 words (46%) in the CDI have fricative-onset consonants. Since the intention was to test children on various possible onsets, an even spread among the manner of articulation of word-onsets was ensured, instead of directly following the proportion in the language. These are henceforth referred to as manner categories of consonant sequences.

Next, the syllable structure of the disyllabic nonwords was determined. Syllable structure refers to the components that make up a syllable. In Mandarin, each syllable is mostly consonant-vowel (CV), with the only CVC shapes ending with nasal /n/ or /ŋ/. From the CDI, the 372 disyllables yielded the following proportions of possible syllable structures: open

syllables in both syllables (S1 and S2; 188 words, 51%), nasal coda in S2 (77 words, 21%), nasal coda in S1 (67 words, 18%) and nasal codas in both syllables (40 words, 11%). The manner categories of the segmental sequences in combination with the syllable structure of the disyllabic words in the CDI are presented in Table 3.1.

Table 3.1: Frequency distribution of manners of articulation of disyllabic words in CDI

Legend: P = plosive, N = nasal, F = fricative, L = liquid, C = consonant, V = vowel, monophthongs/ diphthongs/triphthongs are all represented by a singular V. Tables are ordered in terms of highest to lowest of the manner categories of segmental sequences in the respective syllable structures.

(C)V (C)V		(C)V (C)VN		(C)VN (C)V		(C)VN (C)VN	
CV CV		**CV CVN**		**CVN CV**		**CVN CVN**	
FV-FV	48	FV-FVN	20	FVN-FV	18	FVN-FVN	6
PV-FV	24	FV-PVN	9	FVN-PV	11	FVN-PVN	5
PV-PV	18	PV-PVN	6	PVN-FV	10	PVN-PVN	4
FV-PV	17	PV-FVN	4	PVN-PV	7	PVN-FVN	3
NV-PV	10	NV-PVN	2	LVN-FV	4	FVN-NVN	2
NV-FV	7	NV-FVN	2	FVN-NV	3	FVN-LVN	2
FV-LV	5	FV-NVN	2	NVN-PV	2	LVN-PVN	2
NV-NV	4	PV-NVN	1	PVN-NV	1	PVN-NVN	1
FV-NV	4	PV-LVN	1	FVN-LV	1	PVN-LVN	1
LV-PV	4	NV-LVN	1	LVN-LV	1	NVN-PVN	1
LV-FV	4	LV-NVN	1			NVN-LVN	1
PV-LV	2	LV-LVN	1			LVN-FVN	1
NV-LV	2						
PV-NV	1						
CV V		**CV VN**		**CVN V**		**CVN VN**	
PV-V	6	FV-VN	10	FVN-V	2	PVN-VN	2
FV-V	6	PV-VN	1	PVN-V	1	FVN-VN	1
NV-V	4	NV-VN	1				
LV-LV	1						
V CV		**V CVN**		**VN CV**		**VN CVN**	
V-FV	9	V-FVN	6	VN-PV	2	VN-FVN	7
V-PV	6	V-PVN	3	VN-LV	2		
V-NV	1	V-NVN	2	VN-FV	1		
		V-LVN	1				
V V		**V VN**		**VN V**		**VN VN**	
V-V	5	V-VN	3	VN-V	1	VN-VN	1
total = 372	188		77		67		40
%	0.51		0.21		0.18		0.11

For example, the frequency of fricative-fricative segmental sequence [FV-FV] words like *ɕiɛ4ɕiɛ0* 'thank you' was the most frequently occurring in the language (48 of 372 disyllables, 13%) whereas the frequency of two onset-less vowel-nasal segmental sequence [VN-VN] words like *uan3an1* 'goodnight' was the least frequently occurring (1 of 372, <1%).

Each of these segmental sequences (i.e., [FV-FV], [PV-PV] etc., 74 in total) was ranked from most to least frequently occurring, and a median split (*N* = 6) was used to classify them as frequent or infrequent segmental sequences (Table 3.2).

Table 3.2: Frequent and infrequent segmental sequences in CDI
Legend: P = plosive, F = fricative and affricate, N = nasal, L = liquid, V = vowel.

frequent		infrequent			
FV-FV	48	FV-LV	5	PV-NV	1
PV-FV	24	FVN-PVN	5	PV-NVN	1
FV-FVN	20	V-V	5	PV-LVN	1
PV-PV	18	PV-FVN	4	PV-VN	1
FVN-FV	18	PVN-PVN	4	PVN-NV	1
FV-PV	17	NV-NV	4	PVN-NVN	1
FVN-PV	11	NV-V	4	PVN-LVN	1
PVN-FV	10	LV-PV	4	PVN-V	1
NV-PV	10	LV-FV	4	NV-LVN	1
FV-VN	10	LVN-FV	4	NV-VN	1
FV-PVN	9	FV-NV	4	NVN-PVN	1
V-FV	9	PVN-FVN	3	NVN-LVN	1
PVN-PV	7	FVN-NV	3	LV-NVN	1
NV-FV	7	V-VN	3	LV-V	1
VN-FVN	7	V-PVN	3	LV-LVN	1
PV-PVN	6	PV-LV	2	LVN-FVN	1
PV-V	6	PVN-VN	2	LVN-LV	1
FV-V	6	NV-PVN	2	FVN-LV	1
FVN-FVN	6	NV-FVN	2	FVN-VN	1
V-PV	6	NV-LV	2	V-NV	1
V-FVN	6	NVN-PV	2	V-LVN	1
		LVN-PVN	2	VN-FV	1
		FV-NVN	2	VN-V	1
		FVN-NVN	2	VN-VN	1
		FVN-V	2		
		FVN-LVN	2		
		V-NVN	2		
		VN-PV	2		
		VN-LV	2		

From Table 3.2, the distribution of syllable structures among segmental sequences revealed a confound: syllables with nasal codas tended to be infrequent, and that syllables with fricative-initials tended to be frequent. To ensure that performance on the repetition task was not skewed by syllable structure, an equal number of syllables with and without nasal codas, as well as syllables with fricative-initials was selected across frequent and infrequent segmental sequences. As syllables with nasal codas appeared more in S2 than S1 (mentioned above), this pattern was followed for the nonword stimuli. The final choice of segmental sequences is shown in Table 3.3.

Table 3.3: Selected segmental sequences from CDI for nonword repetition task
Legend: P = plosive, F = fricative and affricate, N = nasal, L = liquid, V = vowel.
The focus of the analyses are the segments coloured in red and blue. **Bold**: referenced in text.

segmental sequences	
frequent	**infrequent**
NV-PV	FV-NV
LV-FV	PV-NV
PV-PVN	PV-FVN
FV-FVN	**LV-PVN**

Unfortunately, a coding error was made early in the study such that syllables that would have been considered diphthongs and triphthongs by the Zhu (2002) notation (which was the intended way of notation for this study, and which would have meant that they would have been classified as vowel-initial) were incorrectly classified as glide-vowel instead. The stimuli had to be recoded after the task had already been administered (the naturalistic and experimental components of the study were run concurrently). This meant that the LV sequences that contained [w] and [j] in the CDI were reclassified as VV, and the 'LV-FV' segmental sequence ended up being considered an 'infrequent sequence' amongst the CDI words (because only /ɹ/ and /l/ remained, both infrequently occurring onsets of words in the CDI) while the 'LV-PVN' segmental sequence ended up being non-existent (because there were no /ɹ/- or /l/-onset words left with this sequence). Reconsidering 'LV-FV' as infrequent and removing the 'non-existent' 'LV-PVN' nonword targets resulted in only 12 remaining frequent segmental sequences and 16 infrequent sequences instead of an equal 16 each. Nonetheless, whether the LV-FV and LV-PVN words were classified as originally planned, or reclassified them following the detection of the error, that ended up making no difference to the findings. The classification was therefore left as it was originally planned, i.e., when LV-FV was counted as 'frequent' and LV-PVN was 'infrequent', as in Table 4.3.

The distribution of tone sequences in the CDI was determined next in order to establish the corresponding tone sequences for the nonword repetition stimuli (Table 3.4). Tone sequences refer to the combination of the two tones of the disyllabic nonword stimuli. This yielded 19 tone sequences. A median split ($N = 19$) was used to classify them as frequent vs. infrequent tone sequences.

Table 3.4: Frequency distribution of tone sequences in CDI

Legend: Percentage (%) is calculated out of 372 disyllables in CDI.

tone sequences					
frequent		%	infrequent		%
T4-4	33	0.09	T2-1	17	0.04
T3-1	29	0.08	T4-3	16	0.04
T1-4	29	0.08	T1-2	15	0.04
T2-3	26	0.07	T2-0	15	0.04
T3-4	24	0.07	T1-3	14	0.04
T1-1	23	0.06	T2-2	14	0.04
T4-1	22	0.06	T1-0	13	0.04
T4-0	22	0.06	T4-2	12	0.03
T2-4	19	0.05	T3-0	10	0.03
T3-2	19	0.05			

Tone combinations that included the neutral T0 were not included for selection, as speakers of Mandarin in Singapore do not tend to neutralise syllable tones (Shang & Zhao, 2013, 2017). The final choice of tone sequences is shown in Table 3.5.

Table 3.5: Selected tone sequences from CDI for nonword repetition task

tone sequences	
frequent	infrequent
T4-4	T1-2
T3-1	T1-3
T1-4	T2-2
T2-3	T4-2

In terms of the selection of the nonword syllables themselves (to fit the manner categories of segmental sequences PV, FV etc.), Wang et al.'s (1963) table of possible Mandarin syllables was consulted first. There were gaps in this table where syllables do not presently exist in the language. The nonword syllables in this study were selected using the gaps in the table by matching existing consonant-onsets with existing nuclei that do not currently occur together. None of the nonword syllables have consonant clusters or illegal codas. Unlike

Wiener and Turnbull (2016), no 'tonotactic accidental gaps'[5] (existing legal syllables that happened not to occur with a particular tone) were included. Wang and Huang (2016) had a nonword *pia1tian1* with an existing second syllable *tian1* 'upside down'; here, none of the nonword syllables had referents. The following exclusion criteria were used for the nonword syllables:

1) Syllables were excluded if they sounded similar to the pronunciations of real words in Mandarin for Singapore speakers (e.g., *foŋ* and *moŋ* were possible nonword syllable candidates but they sound like *fəŋ1* 'wind' and *məŋ2* 'blurry')

2) Syllables were excluded if they contained onsets that children from the pilot study in York (Choo et al., 2019, mentioned in section 2.2.2) produced infrequently (e.g., syllables that began with [f], [ɹ] and aspirated fricatives and affricates). The set of consonant onsets included was narrowed down to: [p, t, tʰ, k, ts, s, h, m, n, l].

3) Syllables that contained the 'y' vowel were excluded due to children's greater difficulty with them as well as their perceptual similarity to 'i' (thus *-y*, *-yɛ*, *-yan* and *-yn* syllables were excluded).

The resultant syllables were then categorised into onsets with plosive, fricative (and affricate), nasal or liquid consonants; they were then combined to form a disyllabic nonword, while controlling for the variables mentioned. As far as possible, each nonword syllable appeared in both syllable positions [first syllable (S1) or second syllable (S2)], and both 'frequent segmental sequence' and 'infrequent segmental sequence' categories. Some syllables appeared only in either the frequent or the infrequent category because there were more instances of that segmental sequence occurring in the respective categories (e.g., there were two frequent sequences with FV but only one infrequent sequence with FV). Even if the segmental sequence had the same consonant manner across both syllables (e.g., the PV-PN and FV-FN segmental sequences), a different consonant was used so there were no nonword targets with consonant harmony (i.e., there were no nonword targets such as *pou3poŋ4* [the same *p*-onset used in the PV-PN sequence] or *tsuai2tsuaŋ1* [the same *ts*-onset used in the FV-FN sequence]. The tone sequences were then added onto the nonword syllable combination.

Frequent and infrequent segmental sequences, and frequent and infrequent tone sequences were then crossed to arrive at four categories of nonword stimuli. The final list of 32 nonword targets for the nonword repetition task is shown in Table 3.6.

---

[5] For example, *su* is a legal syllable, with real words *su1* 'puff pastry', *su2* 'colloquialism' and *su4* 'vegetarian' among others; *su3* would be a tonotactic accidental gap as there is currently no such lexical item.

Table 3.6: List of nonword stimuli

Legend: P = plosive, F = fricative and affricate, N = nasal, L = liquid, V = vowel,
= item number, **bold**: referenced in text.

		#	Frequent segmental sequence		#		
		#	**Frequent tone sequence**		#	**Infrequent tone sequence**	
NV	**P**V	1	*mia4*	***pə4***	9	*mua1*	*tʰiou2*
		2	*nia3*	*tʰei1*	10	*nuai1*	***kia3***
LV	**F**V	3	*lua1*	***hiɛ4***	11	*luei2*	*tsua2*
		4	*luai2*	*sua3*	12	*luei4*	***hiau2***
PV	**P**VN	5	***kiou4***	*poŋ4*	13	*puai1*	*tʰin2*
		6	*pou3*	*tuaŋ1*	14	*tia1*	***kian3***
FV	**F**VN	7	*sei1*	***hin4***	15	*tsuai2*	***hian2***
		8	*suai2*	*tsuaŋ3*	16	***hiɛ4***	*suəŋ2*
			Infrequent segmental sequence				
		#	**Frequent tone sequence**		#	**Infrequent tone sequence**	
FV	**N**V	17	*tsua4*	*mia4*	25	*sei1*	*muai2*
		18	***hiau3***	*mua1*	26	*suai1*	*muei3*
PV	**N**V	19	*tʰiou1*	*nua4*	27	*puai2*	*nia2*
		20	***kia2***	*nuai3*	28	*tia4*	*nuei2*
PV	**F**VN	21	***pə4***	***hin4***	29	***kiou1***	***hian2***
		22	*tʰei3*	*tsuaŋ1*	30	*pou1*	*suəŋ3*
LV	**P**VN	23	*luei1*	*poŋ4*	31	*lua2*	*tʰin2*
		24	*lua2*	*tuaŋ3*	32	*luai4*	***kian2***

Some of these syllables are similar (e.g., *hin* in this study vs. *hio*) or identical (e.g., *kian*, *luai*, *nua*) to what Wang and Huang (2016) had. However, a Mandarin speaker consulted post-hoc about the nonword stimuli expressed concern about the specific co-occurrence patterns of some of these consonants and vowels, also put forth by other authors (e.g., Chen & Kent, 2005; Cheng, 1982; Luo, 1992). These authors note that in Mandarin, alveolar consonants tend to co-occur with front vowels and labials and velars with back vowels, in concert with Davis and MacNeilage's (1990; 1995) frame-and-content theory and Clements' (1991) unified theory of place of articulation for consonants and vowels. Thus, some syllables in the nonword task were made up of combinations where the onset consonant and the first vowel of the nucleus do or do not exist in the language (the latter syllable types are in **bold** in Table 3.6). For example, for the nonword *sei1hin4*, *se* exists in the language (*se4* 'colour'), to which *'i'* was added to make the nonword syllable *sei*. On the other hand, *hi* does not exist in Mandarin presently (it is a glottal consonant with a front vowel, going against the 'alveolar with front vowel' rule), thus the entire *hin* syllable is relatively less wordlike than *sei*. Wang and Huang (2016) also had *hio1kʰian1* as a nonword in their study, combining a glottal consonant with a front vowel. Wang and Huang acknowledged that both children suspected of DLD and the age controls in their study might have repeated the *hio* syllable more poorly

for that reason. However, in the present study, preliminary analyses revealed no differences in the performance of nonword targets containing part-syllables that presently exist or do not exist in the language (the results of a wordlikeness judgement task also revealed no differences in ratings of the nonword stimuli, see later section 3.2.1.1). Since there was no difference, the final set of nonword stimuli analysed is as it appears in Table 3.6.

The first naturalistic play recording session of the experimental study (see section 3.3.2 for the full procedure) was conducted to ensure that children could indeed produce all the ten consonant onsets [p, t, tʰ, k, s, ts, l, m, n, h] present in the nonword repetition task. Following Vihman (2019), consonants were considered 'acquired' if they occurred at least twice in different words, consonants were considered 'marginally acquired' if they occurred only once, and not acquired if they only occurred as substitutions (cf. section 2.2.1). Using Vihman's definition, a majority (17 of 20, 85%) of children had at least marginally acquired the consonants [p, t, k, h, m, n], corresponding with Li and To (2017; previously mentioned in Chapter 2) that [p, t, m, h] were among the first to be acquired. For the remaining consonants [tʰ, ts, s, l], 12 (60%) children had at least marginally acquired them. All the consonants tested were thus those with which the majority of the children had had at least some production experience, even if the children had not fully acquired the consonants yet.

Twenty words that were familiar to 2-year-old children were also selected from the Singapore CDI and added to the repetition task to make the task less daunting (see Appendix A for list of words). The responses to these words were not analysed.

3.2.1.1 Wordlikeness judgement task

To confirm that there were no differences in wordlikeness between the nonword targets with extant part-syllables (i.e., the existent part-syllable *se* in *sei* but non-existent entire syllable *hin* in the example above) that exist in the language and those that do not, a follow-up supplementary wordlikeness judgement study of the nonword stimuli was created and administered to adult native language speakers of Mandarin. Two items from Wang and Huang (2016), *pia tian* and *nua fai*, were used as practice items. All the tone combinations in Wang and Huang (2016) were T1-1 but the intention of the practice items was to emulate the different tone combinations in the actual task, so the T1-1 combination was kept from Wang and Huang for *pia1tian1*, and the T1-4 tone combination, randomly selected from the tone combinations used in the present study, was used for *nua1fai4*.

The 32-item questionnaire (with 2 additional practice items) was hosted online on Qualtrics (https://qualtrics.com) and publicised to anyone who spoke Mandarin as a native language/mother tongue. The ethics committee from the Department of Language and Linguistic Science, University of York, approved the follow-up study before data collection commenced. Thirty participants took part in the study. Participants were told to read through the information provided and select 'yes' to consent before they could proceed. Anyone who selected 'no' to consent was automatically redirected to the end screen. Email addresses were collected to prevent spam. Email addresses of all participants were kept separately from their data: The debrief was sent to these email addresses if participants selected 'yes' to the question asking if they wanted to receive the debrief; email addresses of the remaining participants were discarded before data analysis commenced. All participants heard the two practice items in the same order; the order of the remaining 32 items was randomised across participants.

Participants were asked to rate how much they thought each word sounded like a possible word in Mandarin (could the word be used for a new product or gadget in Mandarin, or does it sound very strange and unlikely as a potential Mandarin word?) on a scale of 1 (sounds extremely unlike a possible word in Mandarin) to 5 (sounds very much like a possible word in Mandarin). They could listen to the audio files as many times as they liked before making a response.

The stacked bar chart rating for each of the four conditions ('freqcombo': FSFT = stimuli with frequent segmental sequences and frequent tone sequences, FSIT = stimuli with frequent segmental sequences and infrequent tone sequences, ISFT = stimuli with infrequent segmental sequences and frequent tone sequences, ISIT = stimuli with infrequent segmental sequences and infrequent tone sequences) are depicted in Figures 3.1 and 3.2, below. The linear regression model to examine the possible interaction of segment frequency and tone frequency on rating is presented in Table 3.7.
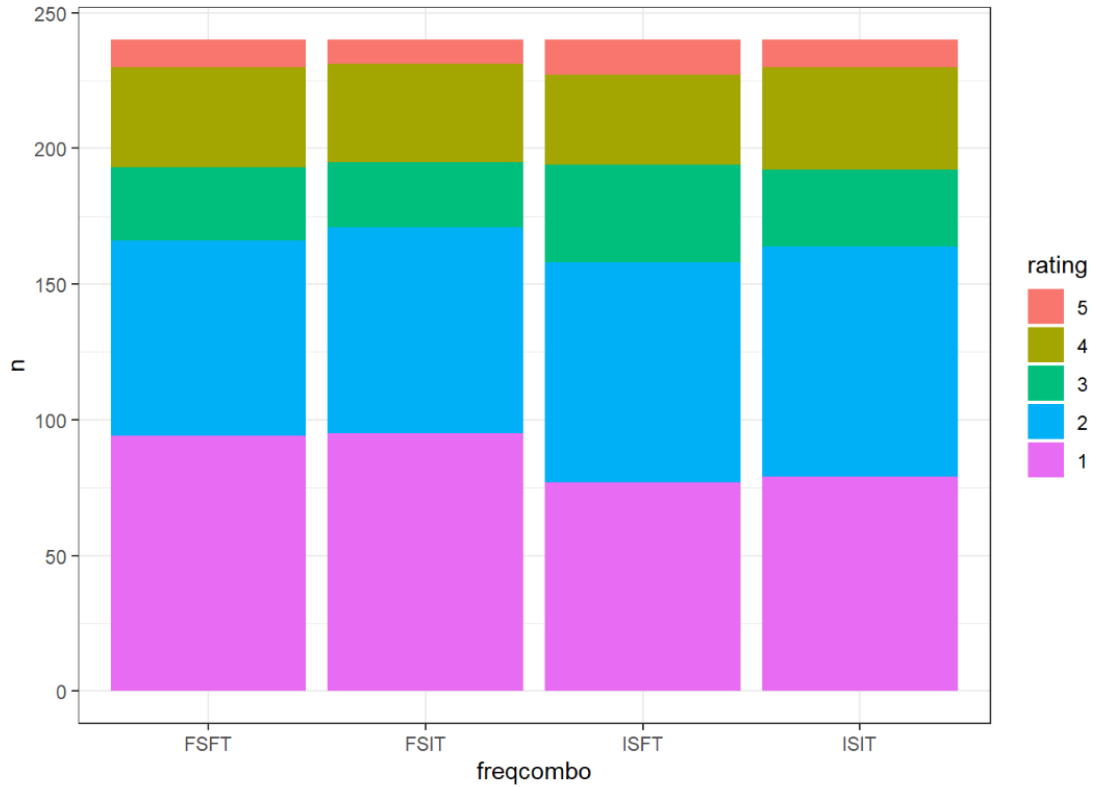
Figure 3.1: Spread of ratings among the four conditions of nonwords (Legend: x-axis represents the four categories of words, y-axis represents the ratings from 30 people × 8 words in each of the four categories, $N = 240$)
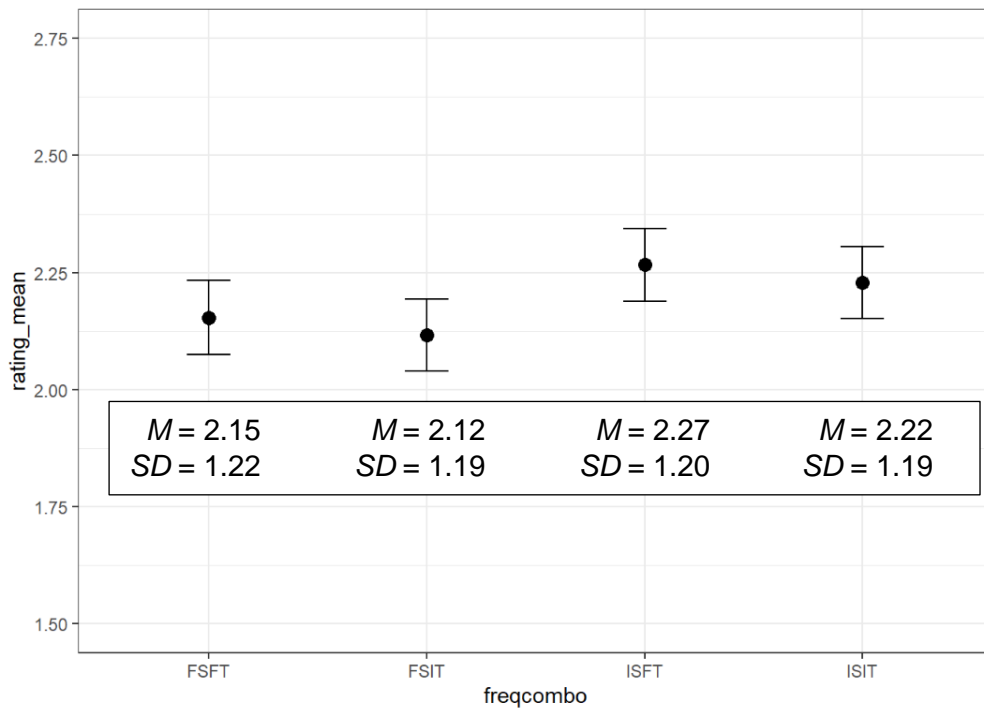
Figure 3.2: Means and standard deviations of each of the four conditions. Error bars denote standard error

The overall mean of the ratings is 2.19 (*SD* = 1.19): the adults rated these nonwords on the lower end of wordlikeness.

Table 3.7: Linear regression model for assessing the effect of segment frequency and tone frequency on ratings

	Estimate	Std.Error	*t* value	Pr(>|*t*|)	
(Intercept)	2.15	0.21	10.50	< .001	***
segfreqIS	0.11	0.23	0.50	0.62	
tonefreqIT	-0.04	0.23	-0.17	0.87	
segfreq:tonefreq	-0.001	0.32	0.00	1.00	

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " "
lmerTest::lmer(rating ~ segfreq * tonefreq + (1 | participant) + (1 | item), data = sample)

The default treatment coding in R was employed, thus the FSFT condition was the reference level (intercept). The rating mean for IS was 0.11 higher than that for FS, meaning that the frequent segmental sequences did not make the words more wordlike, but rather less so, and the rating mean for IT was 0.04 lower than that for FT, meaning that frequent tone sequences may have made words slightly more wordlike. However, as can be seen from the stacked bar charts and the table, there is no significant difference among the four conditions of nonwords. The adults rated the nonwords in each of the four conditions comparably – there was no inherent difference in judged wordlikeness among the nonwords presented in Table 3.6. The adult wordlikeness judgement ratings task confirmed that these nonwords were comparable in terms of (non)wordlikeness.

3.3 Procedure

The recording equipment was a Zoom Q4 camera with Tascam D10L digital audio recorders that recorded onto microSD cards. The videos were set to record in 1080 HD, .mp4 video CODEC file format and the audio was set to record in 44.1khz, 16-bit .wav file format. The child wore the microphone in a cloth vest that was placed unobtrusively as possible to prevent the child from noticing it and handling it; parents clipped the microphone to the front of their shirt. All the recording sessions were then transcribed via ELAN 6.2 (Nijmegen: Max Planck Institute for Psycholinguistics).

3.3.1 Naturalistic study

The children were video- and audio-recorded in their homes in a naturalistic play setting with their mother for half an hour per session for six sessions over six months. The author offered suggestions about reading to the children or playing toys with them but otherwise there was no structure imposed in these sessions; she observed these parent-child interactions and intervened as little as possible. For some children, grandparents were occasionally present but their presence did not influence the recordings.

3.3.2 Experimental study

Two recording sessions were scheduled per child, one week apart, in the month of their second birthday. Parents filled in an informed consent sheet, a language background questionnaire and a Singapore Communicative Development Inventory (CDI; Singh et al., 2022) in the first session.

In the first session, the children were video- and audio-recorded in their homes in a naturalistic play setting with their mother for half an hour. The investigator offered suggestions to the mother about reading to the children or playing with them but otherwise imposed no structure on the sessions; she observed these parent-child interactions and intervened as little as possible. For some children, grandparents were occasionally present but their presence did not influence the recordings. The aim of the first session was twofold: 1) to first familiarise the child with the investigator for maximal compliance during the actual administration of the nonword repetition task in the second session; 2) to ensure that the children could produce all the consonants that were present in the nonword repetition task (see Table 3.6).

In the second recording session, a nonword repetition task was administered to the children. Children heard one of two counterbalanced lists of 32 nonwords (see Table 3.6). The investigator produced the nonword stimuli 'live' for children to repeat. Pre-recorded stimuli were not used so that the investigator could respond to any delay or lack of responsiveness in real time. If children were unwilling to respond, the investigator asked the parents to repeat the form as well, to set an example for the children. The session ended when all 32 nonword targets had been repeated. To facilitate cooperation stickers and verbal encouragement and praise (regardless of accuracy) were peppered throughout the session.

3.3.3 Transcription

3.3.3.1 Segments

Three types of data sources were extracted from the naturalistic study: mothers' speech, child targets and child forms. The relation among the three data sources is: mothers' *input speech* provided a measure of the ambient language the children heard e.g., a mother may be pointing to an elephant in a picture book and saying *ta4ɕiaŋ4* 'elephant', *targets* refer to the adult forms that children were trying to say, e.g., the child responds to her pointing by trying to say *ta4ɕiaŋ4* 'elephant', and *child forms* refer to what children actually produced, e.g., the child produces *ta1taŋ1* for *ta4ɕiaŋ4* 'elephant' in response to the picture book. An elaboration of these terms follows.

Mothers' speech from the first recording session was analysed in order to get a measure of the words children hear from their interlocutors. Transcription of mothers' speech followed Vihman (2019): no more than one variant per word in any one syllable structure was counted (e.g., consonant-vowel consonant-vowel-consonant: CV CVC). That is, types were counted to ensure that there was no over-representation of any one token in a particular session. For each word, as many syllable structures were included per word as the mother's forms warranted. When the focus was on segments, tone changes were disregarded (e.g., if a mother produced 'mother' as *ma1ma0* and *ma2ma0*, it was counted only once, as the forms have the same CV CV syllable structure even though they differ in tone combination; if a mother produced it as *ma1ma0* [CV CV] and *mau2ma0* [CVV CV], it would be counted twice).

Child targets were determined by observing the situational context in the video recordings in which the children produced word forms and their eye gaze/gestures, then analysing them in terms of the ten criteria listed in Vihman and McCune (1994) for identifying words in children's speech (see Table 3.8).

Table 3.8: List of criteria for identifying words (from Vihman & McCune, 1994)

Criteria	Specific conditions	Example: if a child said *mau1* 'cat'
Context-related		
Determinative context	Use of word in a context that strongly suggests that word and no other	Can be observed or inferred from the video recording that there was a picture of a cat in the storybook that mum and child were reading
Maternal identification	Parent identifies at least one instance of the form	Mum agreed and said more about the cat
Multiple uses	Use of word more than once within a time frame	Child used *mau1* when referring to the cat in the storybook or other appropriate contexts
Multiple episodes	Use of word more than once across different time frames	
Vocalisation shape		
Complex match	Child form matches more than two segments of the adult target	*mau1* is a complex, exact and prosodic match (including correct tone) to the adult target
Exact match	Child form matches adult target exactly	
Prosodic match	Tuneful match with adult target and consistent across tokens	
Relation to other variants		
Presence of imitated tokens	At least one instance is imitated	Child imitates mum saying *mau1*
Invariant use	All instances of the word exhibit the same phonological shape	Child used *mau1* when referring to the cat in the storybook or other appropriate contexts
Absence of inappropriate uses	All uses occur in contexts which plausibly suggest the same word	

For child forms, the transcription process followed Vihman (2019) in the same way as the process described above for mothers. Only additional criteria/differences are highlighted. The count of children's words was a cumulative report of the unique number of words the children targeted and produced per session over their five or six recording sessions. Onset consonants were the focus of these analyses (nasal-coda consonants will be mentioned in passing). There were two types of scoring for the segmental element in both naturalistic and experimental studies: a manner of articulation of onset consonant and word structure match. This was to provide an overall idea of the type of responses to targets – whether there was a match in terms of manner of articulation or word structures.

For the manner of onset and word structure match in both naturalistic speech and nonword repetition responses, a percentage precision match was calculated for the percentage of child forms that matched the respective manner categories or word structures of the targets i.e., % of plosive-initial consonants produced as plosives, or % of variegated targets produced as variegated responses. In addition, the scoring criteria used in Keren-Portnoy et al. (2010) was also followed:

1) differences in aspiration were not considered errors because aspiration is not well controlled before 2 years of age (Macken, 1980);
2) phonological processes common at 2 years of age (substitution of /s/ for /ʃ/ or vice versa, deaffrication, affrication of stops, or substitution of a different nasal in the coda position were not considered errors;
3) substitutions of one consonant by another (e.g., [k] by [t]) were not considered errors if the child made this substitution systematically (for the other words they produced in their everyday speech or the nonwords in the task); non-systematic substitutions were considered errors.

Two additional criteria from Keren-Portnoy et al. (2010) were used for scoring the nonwords:
4) if a child produced a segment correctly in its target position for at least one nonword target but substituted it in another nonword target, the substitution was considered an error;
5) complete omission of a consonant (onset or coda) was considered an error.

Vowels were not analysed because vowel productions at the end of the single word period are not yet robust enough to be reliably transcribed (as mentioned in Chapter 1, cf. Davis & MacNeilage, 1995; Davis et al., 2002; Vihman, 2019), so the scoring for the vowels was more lax. For vowels (criteria adapted from Keren-Portnoy et al., 2010),
1) vowels were scored as errors only when there was a radical deviation from the target or when the vowel was omitted. Thus, if the substituted vowel maintained the height and backness of the target vowel or was near the target in the vowel space, this was not considered an error;
2) substitutions among the low vowels, reduction of diphthongs/triphthongs to any subparts, epenthesis of additional vowels, metathesis of vowel diphthongs/ triphthongs, reduction of any vowel to schwa, or substitution of a schwa by any lax vowel were also not considered errors.

Child forms of spontaneous and imitated targets were initially analysed separately because they test different elements of word production: spontaneous targets test children's retrieval of the words from their long-term memory whereas imitated words test their temporary (re)production of a word form that they have just heard an adult (their caregiver) say. However, as the patterns found for the child forms of spontaneous and imitated targets were similar, the counts have been combined for the analyses presented here.

3.3.3.2 Tones

Across both naturalistic speech (both mothers' and children's) and experimental studies, tone was determined by ear and confirmed by acoustic analyses (see section 3.3.4 below for Praat graphs).

Just as tone changes were disregarded when the focus was on segments, segment changes were disregarded when the focus was on tones, (e.g., if a mother produced 'mother' as *ma1ma0* and *me1ma0*, this was counted only as one variant per session, as the forms have the same CV CV syllable structure and same tone combination, even though they may differ by vowel; if a mother produced 'mother' as *ma1ma0* and *ma2ma0*, this was counted as two).

Following Mok et al. (2019, 2020), the criterion for acquisition of tone sequences is accurate production at above-chance level, but with some modifications. In the studies of Mok et al., production tasks involved monosyllabic Cantonese stimuli, so chance level was calculated at 1 of 6 possible Cantonese tones (16.7%). For disyllabic Mandarin stimuli, there are 4 possible tones for the first syllable and 5 possible tones for the second syllable, thus chance is $1/4 \times 1/5 = 1/20$ (5%). The percentage counted in this thesis provides additional information as the number also represents the level of precision (i.e., the degree of match of child form to target) in the use of the tone sequence. For example, a tone sequence may be used frequently (e.g., 70% of child forms), but if many of these instances were to substitute for other tone sequences, the eventual level of precision may be lower (e.g., 20%). Nonetheless, 20% would still mean that a tone sequence has been acquired, but it additionally represents its imprecision in use in substitutions.

3.3.4 Inter-rater reliability

A trained Mandarin phonetician provided a reliability check. Following Mok and Lee (2018), auditory analysis was preferred to acoustic analysis as the recordings from this study were recorded in the children's homes which was subject to background noise (like Mok and Lee's selection of corpus studies, unlike studies by Wong et al., 2005, 2013 which were conducted in experimental settings and so were acoustically 'cleaner'). Mok et al. (2019) also expounded on the balance between the need of reliability testing and that of ensuring that it can be realistically done; they concluded that having two judges transcribe unfiltered stimuli was the most satisfactory option. This principle was adapted: the phonetician transcribed five minutes from the first and last sessions of each of the four children. The first and last sessions were selected so that a wider range of tokens could be analysed from the children – the children may have been shier in the first session and less inclined to speak, whereas they should have spoken more readily in the final session due to familiarity with the investigator and the task. Proportion of agreement on each consonant (disregarding aspiration) for all consonants transcribed by both the investigator and the phonetician was .74 and that for tones was .70.

For the experimental study, 10 test items of six children were randomly selected via an online generator (https://random.org). As filler words were randomly inserted into the task in order to make the task less daunting for children, the transcribed items were a mix of words and nonwords, but the focus was on nonwords. Proportion of agreement on manner of each consonant for all consonants transcribed by both the investigator and the phonetician was .74 and that for tones was .94 for the nonwords (.84 for consonants and .88 for tones for the filler words). This proportion of agreement was not unusual for child transcription as children may not be speaking at a constant volume at all times, and there is background noise in a home environment, resulting in variability in productions and in transcription (cf. variability in neutral tone productions due to acoustically weak productions, Chen & Xu, 2006). For tokens without agreement, the investigator's transcriptions were used as she had spent a larger amount of time with the children and was more familiar with their speech patterns. Nonetheless, acoustic analyses using Praat (Boersma & Weenik, 2012) were also run and will be presented alongside the analyses where relevant.

Chapter 4 – Naturalistic observations: Results and Discussion

The chapter is divided into two parts: segmental (section 4.1) and tonal (section 4.2) production and potential templates in children's word forms in an everyday setting, based on three data sources: (i) the speech children *hear* from their mothers, and the consonant sequences, word structures and tone combinations that children (ii) *target* and (iii) *produce*. The goal here is to examine whether and how the adult language (both the language in general and specific child-directed input) influences the types of consonants, word structures and tone sequences children learning Mandarin produce, identify any differences between Mandarin and other languages, and categorise any consistent patterns or templates that may be common across the children or child-specific.

4.1 Segmental analyses

Segment analyses (specifically, onset consonants) were conducted using three data sources: what children hear, what they target and what they produce. Mothers' input speech provided a measure of the ambient language the children *heard*, targets refer to the adult forms that children were *trying to say*, e.g., *ta4ɕiaŋ4* 'elephant', and child forms refer to what children *produced*, e.g., *ta1taŋ1* for *ta4ɕiaŋ4* 'elephant'.

For this section on segment analyses, the research questions addressed are:

1) Is there overlap among the consonant sequences children hear, target and produce?
2) Is there evidence of templatic usage in the children's segment productions?

4.1.1 Child characteristics and consonant inventories

Figure 4.1 and Table 4.1 below provide information on child characteristics: Figure 4.1 shows the timeline of the sessions; Table 4.1 provides a summary of the number of spontaneous and imitated (in parentheses) targets and child forms in Mandarin and English and their total for each child across the recording sessions. Numbers of words in English are provided but they are not analysed.

Legend: The number above each data point denotes the recording session; the number below each data point denotes the child's age at the time of the recording session. Age is presented in year;month.day. Numbers and ages in parentheses reflect the video recording session for the experimental nonword repetition task, which is not included in this chapter.

Child pseudonym \ Age	1;7	1;8	1;9	1;10	1;11	2;0	2;1	2;2	2;3

Yi Ling
1 — 1;9.6
2 — 1;10.6
3 — 1;11.23
(4), 5 — (2;0.14) 2;0.22
6 — 2;2.9
7 — 2;3.13

Fu Zhen
1 — 1;9.23
2 — 1;11.0
3, (4) — 1;11.28 (2;0.8)
5 — 2;0.25
6 — 2;2.5
7 — 2;3.11

En Ting
1 — 1;7.26
2 — 1;8.18
3 — 1;9.22
4, (5) — 2;0.15 (2;0.22)
6 — 2;2.2

Yan Min
1 — 1;7.14
2 — 1;8.11
3 — 1;9.13
4 — 1;10.25
5, (6) — 2;0.3 (2;0.7)
7 — 2;1.26
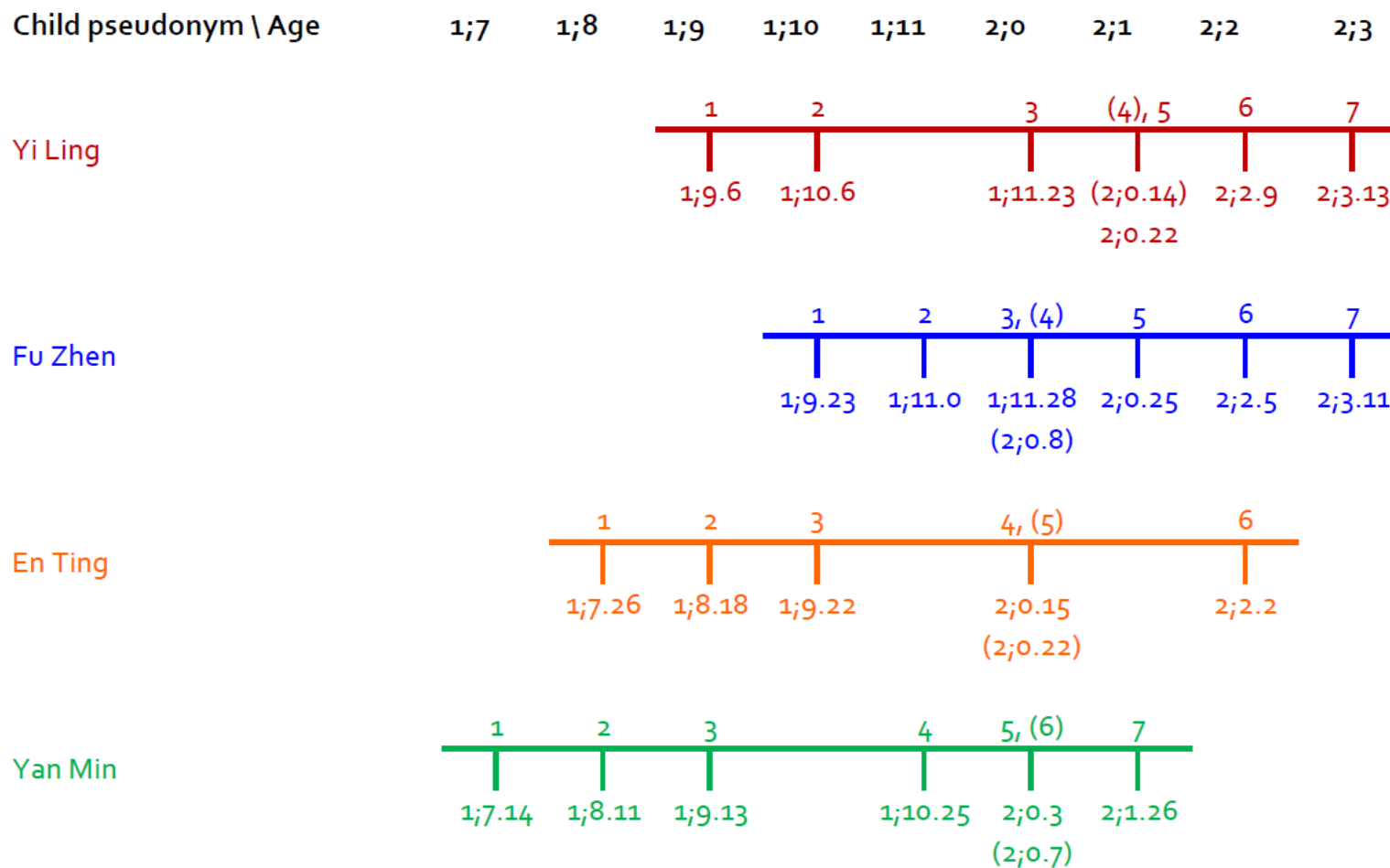
Figure 4.1: Timeline of recording sessions

Table 4.1: Summary table of words produced in Mandarin, English and total for each child across sessions.

Legend: Age is depicted as year;month, numbers in parentheses represent imitations.

Yi Ling	Mandarin	English	Total
1;9	39 (1)	3 (0)	42 (1)
1;10	33 (1)	4 (0)	37 (1)
1;11	36 (8)	1 (0)	37 (8)
2;0	27 (7)	3 (0)	30 (7)
2;2	43 (5)	8 (1)	51 (6)
2;3	35 (9)	2 (3)	37 (12)
Fu Zhen	**Mandarin**	**English**	**Total**
1;10	7 (20)	2 (5)	9 (25)
1;11	5 (24)	4 (0)	9 (24)
2;0	4 (14)	14 (3)	18 (17)
2;1	8 (42)	32 (2)	40 (44)
2;3	9 (44)	35 (7)	44 (51)
2;4	3 (35)	26 (6)	29 (41)
En Ting	**Mandarin**	**English**	**Total**
1;8	6 (10)	40 (0)	46 (10)
1;9	12 (10)	40 (2)	52 (12)
1;10	17 (9)	24 (1)	41 (10)
2;1	8 (3)	42 (3)	50 (6)
2;2	12 (7)	19 (2)	31 (9)
Yan Min	**Mandarin**	**English**	**Total**
1;7	42 (15)	3 (2)	45 (17)
1;8	35 (16)	4 (0)	39 (16)
1;9	46 (28)	4 (0)	50 (28)
1;11	36 (18)	3 (1)	39 (19)
2;0	40 (24)	4 (0)	44 (24)
2;1	47 (16)	3 (1)	50 (17)

Yi Ling and Yan Min were more Mandarin-dominant, Fu Zhen and En Ting were more English-dominant. Although Fu Zhen did not reach the 25wp until the fourth session, there is good reason to believe that she was already there from the first session; her interlocutor – her grandmother, rather than her parents – did not provide as many opportunities to produce words as her mother would have. Although Fu Zhen's interlocutor is her grandmother, the term 'mothers' will be used as shorthand to indicate the adult input that the children heard in the recording sessions.

Table 4.2 shows the children's inventories of consonants. Two consonant counts are provided: the first count establishes the consonants the children had acquired by the first recording session (leftmost two columns) and the second count shows the consonants the children had acquired by the last recording session (rightmost two columns), using Vihman's (2019) criteria for acquisition (mentioned earlier; see also legend of Table 4.2 below).

Table 4.2: Children's consonant inventory in first session and by final session

Legend: Consonants without parentheses = occurred at least twice in different words, considered acquired; ( ) = occurred only once, considered marginally acquired; [ ] = occurred only as substitution for other consonants, not considered acquired (following Vihman, 2019)

Child (first)	First session inventory	By session	Final session inventory

Yi Ling 1;9
N = 11
	p (pʰ) (h)
	t (tɕ)
	(k) ts tsʰ
	[tʂ]
	
	m (n)
		by 2;3
		N = 21
			p (pʰ) f h
			t (tʰ) tɕ tɕʰ ɕ
			k kʰ ts tsʰ s
			            (tʂ) (tʂʰ) ʂ
			
			m n l
			    ŋ

Fu Zhen 1;10
N = 12
	p (pʰ)
	[t] [tʰ] (tɕ) (ɕ)
	k (ts)
	                  (ʂ)
	
	m (n)
	    ŋ
		by 2;4
		N = 19
			p pʰ f h
			t tʰ (tɕ) tɕʰ (ɕ)
			k kʰ ts (tsʰ) (s)
			                         (ʂ)
			
			m (n) (l) (ɹ)
			    ŋ

En Ting 1;8
N = 10
	p [f] (h)
	(t) (tʰ)
	k [ts] [s]
	
	m [l]
		by 2;2
		N = 20
			p [pʰ] [f] h
			t (tʰ) (tɕ) [tɕʰ] ɕ
			k (kʰ) (ts) (tsʰ) [s]
			            [tʂ] (tʂʰ)
			
			m (n) (l)
			    ŋ

Yan Min 1;7 N = 9	p h t (tɕ) (ɕ) k [ts] m n	by 2;1 N = 20	p (pʰ) (f) h t tʰ tɕ (tɕʰ) ɕ k kʰ (ts) (tsʰ) s tʂ [tʂʰ] ʂ m n l

From Table 4.2, the consonants acquired by all four children at the first recording session were /p, m/; by the last recording session, additional plosives /t, k/ and fricatives /ɕ, h/ had also been acquired. The fricatives /tɕ, tɕʰ, ts, s/ were also marginally acquired by the last session. The latter patterns of acquisition of fricatives match the findings of previous work on fricatives (Li & Munson, 2016; Ma et al., 2022); the general trend of acquisition also matched Li and To's (2017) findings: /p, t, m, h/ were all acquired by the children in this study, there was only emerging evidence of acquisition of the retroflex consonants /tʂ, tʂʰ, ʂ/ by the last session, while /ɻ/ was not acquired.

The most frequently produced word-initial consonants used across all children are presented in Figure 4.2.
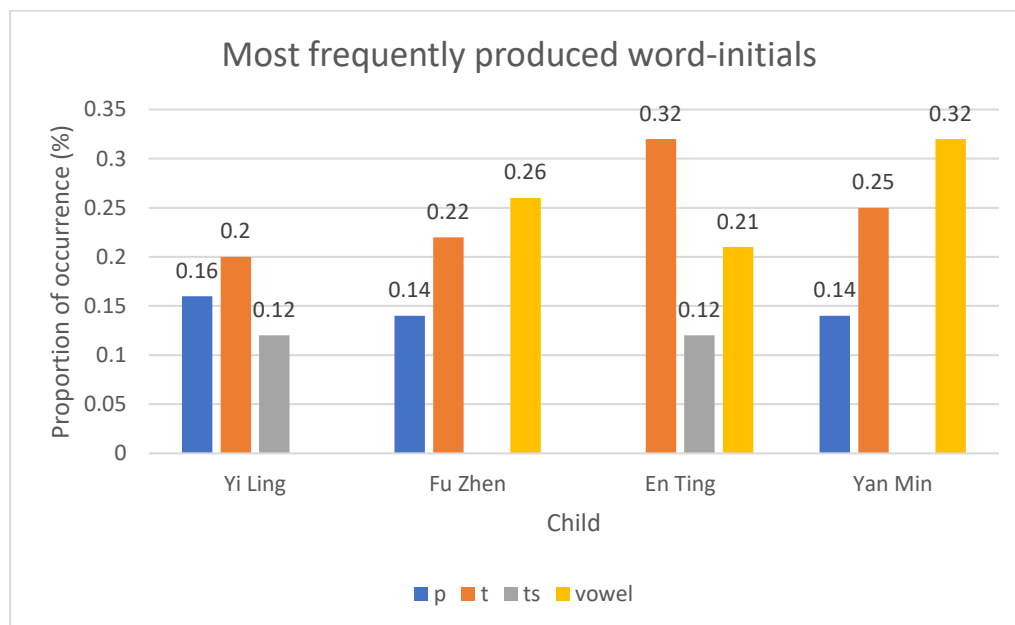
Figure 4.2: Most frequently produced word-initial consonants

The segments [p, t, ts, vowel] were the most frequently used word-initials for all four children, with differing proportions among the children. Yi Ling and En Ting favoured [t], while Fu Zhen and Yan Min favoured vowel-initial words. Yan Min's real name starts with a vowel, so this might have primed her to produce more vowel-initial words. En Ting mostly used her high proportion of [t] in substitution for fricatives. The prominent use of /t/ here will show up in the next analysis on consonant sequences.

4.1.2 Consonant sequences in disyllabic words

The focus of this section is on the consonant sequences of children's responses. The proportion of consonant sequences across the three data sources was examined: Figure 4.3 provides a summary of the distribution in manner categories of the words from mothers' input, the words the children targeted and the child forms produced over the sessions. The top ten most frequent consonant sequences from each data source are presented for each child. [For reference, p-p indicates a /p/-/p/ sequence (e.g., *pa4pa0* 'father') etc.; since the focus is on consonants, vowels are grouped together as 'v'.]

The most frequently occurring consonant sequence across all children will be discussed first, before each individual child's favoured consonant sequences are considered, to identify any templatic usage.
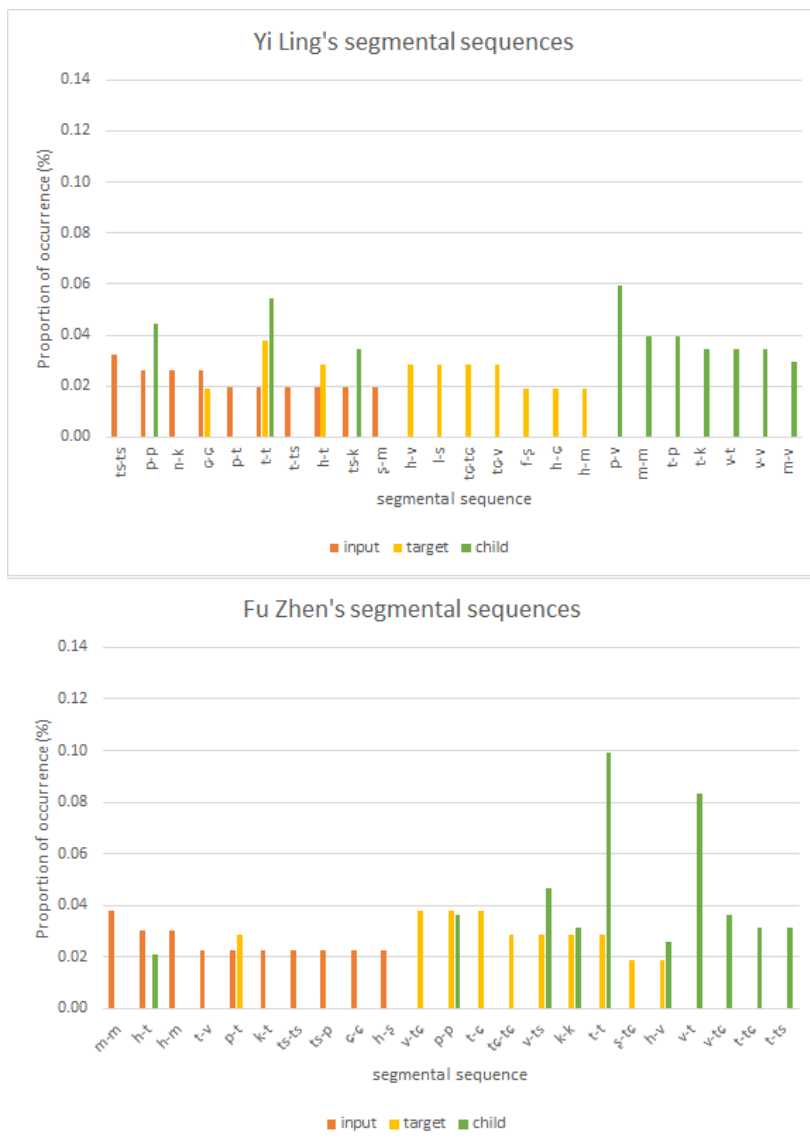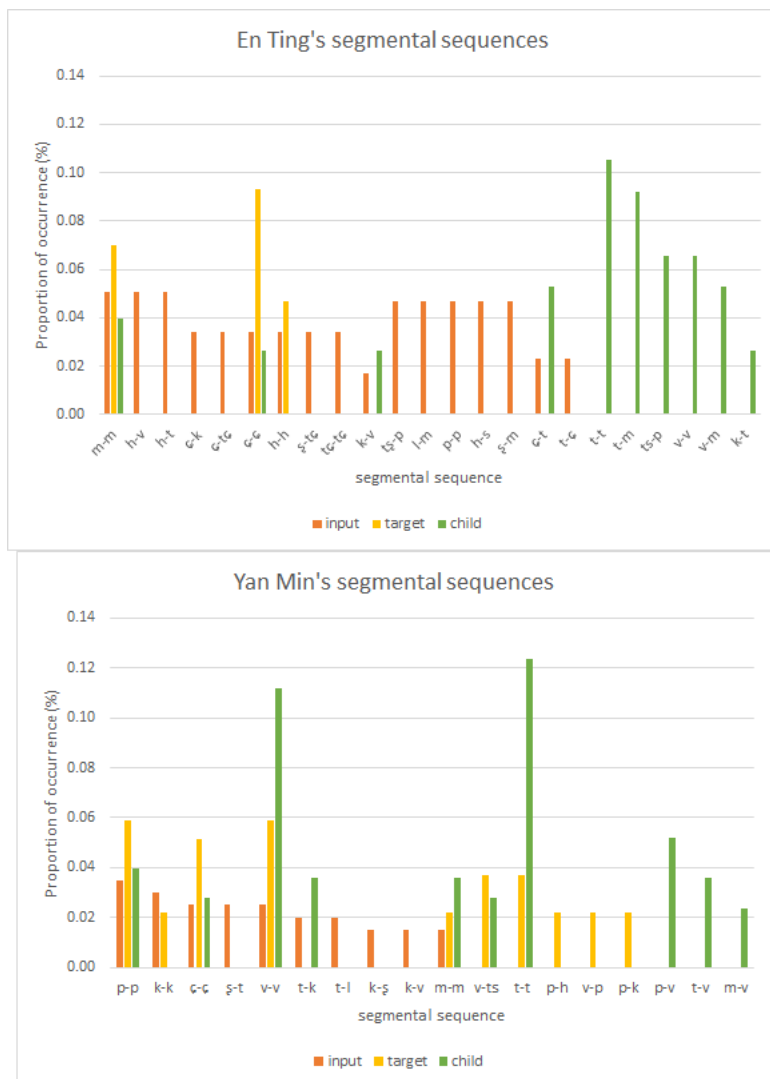
Figure 4.3: Most frequent consonant sequences by child

From Figure 4.3, it is evident that the amount of overlap between mothers' input, the words children targeted and the child forms varies by child. Yan Min's most frequently occurring consonant sequences showed the most overlap over the three data sources; for the other children, it seemed that children targeted and produced word forms more or less independently of what they are hearing from their mothers.

In general, /t/-/t/ is the consonant sequence produced the most frequently by all the children, but its occurrence in the data sources differs by child:, it was present in all three data sources for Yi Ling, only in the child forms for En Ting, and in both the words targeted produced for Fu Zhen and Yan Min.

The most frequently produced consonant sequences and some examples of the selections and/or adaptations to the forms are presented in Table 4.3.

Table 4.3: Selections and adaptations to the most frequently occurring consonant sequences for each child

Legend: coloured text – referenced in text

child	consonant sequence	**selection** target	child form	**adaptation** target	child form	comment
Yi Ling	/t/-/t/	*tiɛ2tau3* 'fall down'	*tiɛ2tau3*	*tʂən3tʰou2* 'pillow'	*tɛn3to2*	consonant harmony
		ti4ti0 'younger brother'	*ti1ti0*	*tʂu1tʂu1* 'piggy'	*tu1tu1*	deaffrication/consonant repetition
		tʰoŋ4tʰoŋ4 'painful'	*tʰoŋ4tʰoŋ4*	*tsɹ4tɕi3* 'by myself'	*ti1ti0*	
	/p/-/v/	*pu2iau4* 'don't want'	*pi1iau4, pu1iau4*	none		
	/p/-/p/	*pa4pa0* 'father'	*pa4pa0* 'father'	*tsuei3pa1* 'mouth'	*puei1pa1*	consonant harmony
		pʰo3pʰo2 'grandma'	*pʰo3pʰo2* 'grandma'			
Fu Zhen	/t/-/t/	*tiɛ2tau3* 'fall down'	*te2tau3*	*toŋ1ɕi0* 'thing'	*toŋ1ti1*	consonant harmony
		tau3tan4 'mischievous'	*tau3tan4*	*tɕiɛ3tɕiɛ1* 'older sister'	*tɛ1tɛ1*	deaffrication/consonant repetition
		tʰoŋ4tʰoŋ4 'painful'	*toŋ1tʰoŋ4*	*tɕi1tan4* 'egg'	*ti1tan0*	deaffrication/consonant harmony
				ʂuei4tɕiau4 'sleep'	*tui1tiau4*	
	/v/-/t/	*ə3tuo1* 'ear'	*ə3tuɒ1*	*ua4tsɹ0* 'socks'	*ua4tɻ3*	deaffrication
		ian3tɕiŋ1 'eye'	*en3tiŋ1*	*ia2tʂʰ3* 'teeth'	*ia2tɻ3*	
		i1tɕia4 'clothes rack'	*i1tɕia4*	*aŋ1tsaŋ1* 'dirty'	*an1taŋ4*	
				tɕien3tau1 'scissors'	*en3tau1*	omission of affricate
	/v/-/ts/	*ua4tsɹ0* 'socks'	*ua4tsɹ3*	*ɕiɛ2tsɹ0* 'shoes'	*ɛ2tsɹ3*	deaffrication
		i2tsɹ3	*i3tsɹ0* 'chair'	*tɕien2tʂɹ3* 'cut paper'	*en3tsɹ3*	omission of /k/ but produced /k/ elsewhere
				kuo1tsɹ0 'pot'	*uo1tsi0*	

77

child	consonant sequence	**selection**		**adaptation**		comment
		target	child form	target	child form	
En Ting	/t/-/t/	*ta4tə0* 'big one'	*ta4tə0*	*ɕiau3tə0* 'small'	*tiau3tə4*	consonant harmony but /ɕ/ used correctly elsewhere: *ɕiau3tə4* and *ta1ɕiaŋ4* respectively
				ta4ɕiaŋ4 'elephant'	*ta1taŋ1*	
				tɕiɛ3tɕiɛ1 'older sister'	*tɛ3tɛ1*	deaffrication/consonant repetition
	/t/-/m/	none		*tsʰau3mei2* 'strawberry'	*tau3mei2*	only used as substitutions for fricatives/affricate and one instance of /l/
				ʂən3mə0 'what'	*tu1mɔ4*	
				ɕia4miɛn4 'below'	*ta1meŋ4*	
				ʂaŋ4miɛn4 'above'	*ta1men4*	
				li3mien4 'inside'	*tau4mən2*	
	/ts/-/p/	none		*tʂə4piɛn1* 'here'	*tsə1pe1*	substitution of /tʂ/ in four of five instances of *tʂə4piɛn1* 'here'
				ʂəŋ1piŋ4 'fell sick'	*tsʰən1piŋ4*	substitution of /ʂ/
	/v/-/v/	*ai3io2* 'oh dear'	*ai3io2*	*kuo4lai2* '[bring it] here'	*u3ai2*	omission of /k/
				hau3hə1 'delicious (for drinks)'	*ə3ə2*	omission of /h/
				hai2iou3 'still have'	*ai1io3*	
				hei1hei1 'black black'	*ei1ei1*	

child	consonant sequence	selection		adaptation		comment
		target	child form	target	child form	
Yan Min	/t/-/t/	tʰou2tʰou2 'head head'	tou2tou2	ɕie4ɕie0 'thank you'	te1ɕie4	substitution of /ɕ/ but used correctly in another instance of 'thank you': ɕie4ɕie4
		tiɛ1tau3 'fall down'	te1tau0	ɕiŋ1ɕiŋ0 'star'	ti1ti1	substitution of /ɕ/ and consonant repetition
				tɕiɛ3tɕiɛ1 'older sister'	te3te2	deaffrication/consonant repetition
				tʂən3tʰou2 'pillow',	tə3to2	deaffrication/consonant harmony
				kʰou3tai4 'pocket'	to3tai1	fronting/consonant harmony but /k/ used correctly in the example for 'this'
				tʂə4kə0 'this'	tə1kə4	deaffrication
	/v/-/v/	a1i2 'aunty'	a3i2	pu4hau3 'no'	au1ua0	omission of onsets
		iaŋ2iaŋ3 'itchy'	ia2ia3	tsʰai3hoŋ2 'rainbow'	ia3o2	
		uan3an1 'goodnight'	ua3an1	ɕiau3ɕioŋ2 'little bear'	ə2ɒ1	
				hə1ʂuei3 'drink water'	ə2so1	
				hua4hua4 'draw'	ua1ua1	
				yɛ4liaŋ4 'moon'	e1ia4	
	/p/-/v/	pu2iau4 'don't want'	pi2au0, pu2iau4	pi4hu3 'lizard'	pi1u0	omission of onsets
				pʰo4lə0 'torn'	pau1uə0	
				pu2huei4 'don't know how to [do something]'	pu2uei4	

Fu Zhen produced /ts/ accurately in some instances and substituted it with /t/ in other instances (e.g., in 'socks' – see red text in Table 4.3). The two /v/-/t/ and /v/-/ts/ consonant sequences are in the V CV vowel-initial forms, which she produced the most frequently (see earlier Figure 4.2). It is also noteworthy that the *tsɿ0* syllable (in blue) in many of her /v/-/ts/ words is a bound morpheme that forms a part of many nouns, so the high frequency of occurrence and practice may have supported production.

The accurate production of /ɕ/ in some words before the accurate production of other voiceless sibilant fricatives, despite its low frequency (seen here in En Ting and Yan Min) has been reported in other studies of Mandarin (Li & Munson, 2016), as well as in other languages (Japanese: Nakanishi, Owada, & Fujita, 1972; Russian: Zharkova, 2005, Polish: Ingram, 1988b, all cited by Li & Munson, 2016). This early production has been attributed to its articulatory ease relative to the other fricatives: producing /ɕ/ uses the larger and more central muscles of the tongue, whereas producing /s/ or /ʃ/ requires smaller and more peripheral muscles of the tongue tip (cf. Davis & MacNeilage, 1990; 1995). The fricative is also personally relevant for Yan Min, given that it occurs in her real name; she also produced it accurately in another frequently occurring word, 'thank you'.

As mentioned earlier, templates may be identified by selection and adaptation to a particular form, and may be indicated by high frequency of use alongside only moderate precision. While /t/-/t/ was used most frequently by all the children, it was used mostly as substitutions for consonants that they have not yet acquired, even by the last session of recording (based on their respective consonant inventories in Table 4.2). As mentioned in Chapter 1, mere substitutions for sounds not acquired yet are not considered templatic.

Li and To (2017) found that initial consonants tended to be deleted, substituted or harmonised (termed 'assimilated' in their review) and vowels tended to be reduced or substituted. In other studies, the deletion of the final consonant is also common in child productions (cf. Macken, 1979/2013 in Spanish, Demuth & Johnson, 2003 in French). Additionally, the most frequent phonological processes Li and To found for consonant substitutions were deretroflexion (Zhu & Dodd, 2000 termed this 'fronting'), backing and stopping, followed by affrication and deaspiration (cf. Cho, 2008; Hsiao, 2008; Jeng, 2011). Most of these processes were found in the present data except affrication – none of the children produced an affricate where a plosive was expected.

4.1.3 Children's responses to disyllabic words with various word structures

In the previous section, the focus was on the consonant sequences that children heard, targeted and produced. In this section, the word structures of the children's responses as a whole are explored.

Children's words were first classified into five word structures (see Table 4.4). These categories of word structures, adopted from Vihman (1978, 2019), were originally defined to prioritise reduplication and consonant harmony, given that these were the two substitutions frequently reported as responses to variegation in other languages (e.g., Stoel-Gammon & Stemberger, 1994; Levelt, 1994; Vihman, 1978; Vihman, 2019). Thus 'other' or 'vocalic' words that were reduplicated or harmonised were grouped under the reduplicated or harmonised labels (e.g., *iaŋ2iaŋ3* 'itchy' was considered a reduplication).

The findings from the previous section showed that children's substitution patterns tended to include few instances of consonant harmony, unlike what has been reported for other languages. These categories of word structures were used to classify both words that children targeted and the forms that the children produced. As a result, one more important distinction needs to be made: the term 'consonant repetition' (relatedly, 'consonant-repeated' forms') will be used when describing the internal characteristics of the word targeted and/or produced by adults and children alike [contains the same two consonants in a word, e.g., *mu4ma3* 'wooden horse'], and the term 'consonant harmony' (and 'harmonised forms') will be used when discussing children adopting it as a substitution pattern or 'compensatory strategy' [making one of the consonants the same as the other, e.g., *ta1taŋ1* for *ta4ɕiaŋ4* 'elephant'] (cf. Menn, 1983; O'Seaghda et al., 2013; see examples in Table 4.5). The goal here was to examine if there were consistent patterns of word structures children hear from their mothers, target and produce and whether there were any consistent patterns of substitutions that may be child-specific or common across all children.

Table 4.4: Definitions and examples of word structures

word structures	definition	example
reduplication (RED)	repetition of a syllable in succession regardless of tone	*ma1ma0* 'mother'
consonant repetition (CR, used for target form) or consonant harmony (CH, used for substitutions)	use of the same supraglottal consonant onset for consecutive syllables regardless of vowel	CR: *mu4ma3* 'wooden horse' CH: *ta1tan1 for ta4ɕiaŋ4* 'elephant'
variegation (VAR)	use of different supraglottal consonant onsets across consecutive syllables	*tʂə4kə0* 'this'
other	words including only one supraglottal consonant onset, with the other being a vowel or liquid, i.e., - V CV or CV V - LV CV or CV LV	*y4mi3* 'corn' *tʂʰu1lai2* '[something has come] out'
vocalic	words with two zero-onset syllables, i.e., - V V - V VC or - VC VC	*a1i2 'aunt'* *y2san3* 'umbrella' *uan3an1* 'goodnight'

Table 4.5 provides a summary of the distribution in word structures of the words the children targeted and produced over the sessions. Mothers' input and targets are presented in the left block of the table, and the word structures in the child forms are presented in the right block of the table. A final summary of all mothers' and all children's data are presented at the end.

Table 4.5: Distribution of word structures of words targeted and produced by all children

Legend: RED = reduplication, CR = consonant repetition/CH = consonant harmony, VAR = variegation. *N* = number of child forms. Percentage (%) refers to the percentage of number of child forms produced as the respective word structures, calculated out of the total number child forms (e.g., number of reduplicated words out of total number of child forms, number of variegated words out of total number of child forms etc.). pre = precision refers to the % of child forms that matched the respective word structures of the targets (e.g., % of reduplicated targets produced as reduplicated, % of variegated targets produced as variegated etc.). Underlined text indicates highest proportion each in input, targets and child forms. **Bold and underlined text**: referenced in text

		input	targets		child forms				
					RED	CR/CH	VAR	other	vocalic
Yi Ling	RED	0.09	0.14		0.83	0.07	0.07	0.03	0
(*N* = 202)	CM	0.04	0.01		0	1	0	0	0
	VAR	0.62	0.55		0.01	0.03	0.87	0.09	0
	other	0.25	0.28		0	0	0	0.91	0.09
	vocalic	0	0.01		0	0	0	0	1
				%	0.12	0.04	0.49	0.31	0.03
				pre	0.96	0.38	0.98	0.83	0.29
Fu Zhen	RED	0.08	0.15		0.72	0.03	0.17	0.07	0
(*N* = 192)	CM	0.07	0.04		0	0.86	0	0.14	0
	VAR	0.52	0.44		0	0.13	0.65	0.21	0
	other	0.32	0.38		0	0.01	0.03	0.94	0.01
	vocalic	0.02	0.00		0	0	0	0	0
				%	0.11	0.10	0.32	**0.46**	0.01
				pre	1.00	0.32	0.89	0.76	0
En Ting	RED	0.07	0.17		0.69	0.08	0.08	0.15	0
(*N* = 76)	CM	0.07	0.07		0	1	0	0	0
	VAR	0.55	0.57		0	0.09	0.70	0.19	0.02
	other	0.29	0.20		0.07	0	0.07	0.67	0.2
	vocalic	0.02	0		0	0	0	0	0
				%	0.13	0.13	0.42	0.26	0.05
				pre	0.90	0.50	0.94	0.50	0
Yan Min	RED	0.10	0.20		0.55	0.18	0.10	0.16	0.02
(*N* = 252)	CM	0.06	0.07		0	0.65	0.18	0.12	0.06
	VAR	0.59	0.37		0	0.15	0.65	0.18	0.02
	other	0.22	0.29		0	0.01	0.04	0.82	0.13
	vocalic	0.02	0.07		0	0	0.06	0.11	0.83
				%	0.11	0.14	0.29	**0.35**	0.11
				pre	1	0.31	0.83	0.67	0.54

		all input	all targets		all child forms				
					RED	CR/ CH	VAR	other	vocalic
all	RED	0.09	0.17		0.67	0.11	0.11	0.11	0.01
children	CM	0.06	0.05		0	0.78	0.09	0.09	0.03
(N = 722)	VAR	_0.57_	_0.44_		0	0.10	0.73	0.16	0.01
	other	0.26	0.31		0	0.01	0.03	0.88	0.08
	vocalic	0.01	0.03		0	0	0.05	0.10	0.85
				%	0.12	0.10	_0.37_	_0.36_	0.06
				pre	0.98	0.35	0.91	0.73	0.45

In general, for all the children, there was a match in proportion across all three data sources (i.e., mothers' input and children's targets in left block of table and child forms in right block of table): the highest proportion for all three data sources was variegated words, followed by 'other'. Using the Singapore CDI as a measure of frequency in the ambient language, 227 (61%) of 372 words are variegated, followed by 114 (31%) 'other', 18 (5%) are reduplicated and 13 (3%) are consonant-repeated. The children's word forms followed a similar proportion of the word structures in the input and in the ambient language. Nonetheless, there were some differences among the children: Yi Ling and En Ting produced variegated words most frequently, whereas Fu Zhen and Yan Min produced 'other' words most frequently (see **bold and underlined** text).

For word structures, unlike the earlier analysis of the word-initial segment, the *most frequently produced* word structures also had the highest level of *precision* (i.e., the degree of match of the word structure of the child form to the target: producing reduplicated words for reduplicated targets, variegated words for variegated targets etc.). Reduplicated and variegated words had the highest level of precision (above 90%), followed by 'other' words. The precision for reduplicated words is high, likely because many of the reduplicated words produced are kinship terms (see discussion in section 4.2.3.2). Variegated words tend to become reduplicated or harmonised in other languages (cf. Stoel-Gammon & Stemberger, 1994; Vihman et al., 2023), but children in the present study were highly precise in maintaining variegation for variegated targets as well (see discussion in section 4.1.4.2). If there were substitutions, the pattern was 'other' rather than reduplication or harmony. This is why 'other' words have only moderate precision: even though 'other' targets were produced as 'other' most frequently (see the high proportion of 'other' responses under each row of 'other' targets), the 'other' word structure was also the next highest substitution for variegated words (see the proportion of substitutions under each row of variegated targets), leading to a drop in precision. Meanwhile, vocalic and forms had the lowest precision. The low precision for both of these forms is understandable because there were not many vocalic

or consonant-repeated targets to begin with, so vocalic and consonant-repeated forms were used mostly only as substitutions.

4.1.4 Discussion

The focus of this chapter was the patterns of segment production and word structures in children's speech and identification of any templates in the children's segment productions. There were matches and mismatches between the ambient language effects and child productions, as well as some findings that contrasted with Lou (2020), which might be due to differences in language environments, lexical advance and/or cultural differences. Additionally, there was no evidence of segmental templatic usage in this study. A summary of the findings from the present study is first discussed, followed by the implications of these findings.

4.1.4.1 Higher proportion of fricatives in Mandarin input relative to other languages

For segment analyses, the rate of overlap and difference among mothers' input, children's targets and child forms of consonants, and how children made substitutions was examined. In brief, in terms of consonants, there was a relatively high proportion of fricatives in mothers' input and children's targets (between 40 and 50%), unlike what has been found in previous studies. Fricatives have not been found to occur frequently in early child-directed speech in studies of other languages (like English, French or Swedish, cf. Vihman et al., 1994), unlike Mandarin, where the count of fricatives (and affricates) in the consonant inventory of the language is substantial: there are 11, making up 52% of the consonant inventory (with 28% plosives and 10% each of nasals and liquids). The later acquisition of fricatives typically found in children learning other languages may thus be due to the infrequent occurrence of these consonants in those languages.

For comparison, the data from Vihman et al. (1994) are presented in Table 4.6 alongside Mandarin data from Lou (2020) and the present study. Vihman et al. (1994) studied the speech of mothers to their one-year-old children, five each learning three languages monolingually: English, French and Swedish. The child forms were cumulative over 9 months while one session each of the mothers' speech was recorded when the children were 12 or 13 months old.

Table 4.6: Proportion of manner categories of mothers' input and child forms in various languages.

Legend: Underlined text indicates highest proportion in each language. Data from English, French and Swedish are taken from Vihman et al. (1994). Data from Mandarin are from Lou (2020) and the present study (grey cells).

Mothers' input	N children	plosive	fricative	nasal	liquid
English	5	0.55	0.12	0.15	0.17
French	5	0.48	0.12	0.21	0.19
Swedish	5	0.58	0.09	0.18	0.15
Mandarin (Lou, 2020)	5	0.33	0.39	0.12	0.16
Mandarin (present study)	4	0.30	0.53	0.13	0.04
mean		*0.45*	*0.25*	*0.16*	*0.14*

child forms	N children	plosive	fricative	nasal	liquid
English	5	0.75	0.07	0.13	0.04
French	5	0.61	0.07	0.22	0.10
Swedish	5	0.78	0.03	0.14	0.05
Mandarin (Lou, 2020)	5	0.43	0.23	0.18	0.16
Mandarin (present study)	4	0.57	0.29	0.11	0.03
mean		*0.63*	*0.14*	*0.16*	*0.08*

From Table 4.6, plosives dominated both the mothers' targets and child forms in English, French and Swedish. While plosives make up the highest proportions in both input and output in these three languages, children are nevertheless producing a higher proportion of plosives than they are hearing in the input.

For the two Mandarin sources, however, there is a difference between what children are hearing from their mothers and what children are producing. While mothers in Lou (2020) produced plosives and fricatives fairly equivalently (30-40%), mothers in the present study produced almost twice the proportion of fricatives (53%) as that of plosives (30%). Nonetheless, while there may be a higher proportion of fricatives than plosives in what their mothers produced, the children in both Mandarin datasets still produced plosives the most frequently, like the children learning the other three languages.

On the other hand, the proportion of fricatives produced by the Mandarin children is higher than that in the other languages, relative to the other manner categories: fricatives are produced about half as much as plosives for Mandarin (on average, about 25% fricatives vs. about 50% plosives), whereas fricatives make up only about 10% of the plosives produced for the other three languages. The frequency of occurrence of consonants in children's speech may be affected by the proportion of occurrence in the ambient languages, but the

frequency distribution of the input may not necessarily be directly mirrored in children's output. A further analysis of the plosive-fricative discrepancy shows that children actually did target fricative-initial words the most (thus reflecting the proportion in the ambient language), but they substituted them with plosive-initial words, resulting in the increase in plosive-initial output relative to the input. This likely reflects the articulatory ease of plosives relative to fricatives (cf. MacNeilage & Davis, 1990). Input frequency is never the whole story (a thread continued in Chapter 5; cf. Menn, 2013, Ota, 2013); the production of segments may also be dependent on articulatory ease. Li and To (2017) additionally found affrication and harmony to be frequent patterns of substitutions, but these were not present in the current study. This leads to the next sub-section on word structures.

4.1.4.2 Presence of more variegation and less reduplication and harmony in Mandarin child forms

The next question addressed related to examining the rate of overlap and difference among mothers' input, children's targets and child forms of the different word structures of words, and how children tended to make substitutions. In brief, there was a match among all three data sources: mothers' input, children's targets and child forms were mostly of the variegated or 'other' categories.

Children tended to be quite precise in producing the word structures of the words they were targeting, i.e., reduplicated targets tended to be produced as reduplicated child forms, variegated targets tended to be produced as variegated child forms, etc. Furthermore, variegated targets were still produced as variegated responses, even if the children substituted either or both consonants, unlike studies of other languages where harmony was the substitution of choice for variegated targets (e.g., Stoel-Gammon & Stemberger, 1994; Vihman, 1978; Vihman, 2019). Reduplicated and harmonised forms are said to be easier because the consonant is held constant. This is why throughout the study of other languages (cf. Kehoe, 2015; Vihman et al., 2022), the proportion of reduplications and consonant harmony in children's productions is usually higher than those of variegated words. Nonetheless, the children produced variegated responses for variegated words, and sometimes even changed the reduplicated or harmonised words into variegated forms (e.g., Fu Zhen and Yan Min). The discrepancy between variegated and harmonised forms for variegated targets is further analysed in what follows.

For comparison, data from Vihman et al. (2022) are presented in Table 4.7 alongside Mandarin data from Lou (2020) and the present study. Vihman et al. examined the extent of child variegated responses produced for variegated targets, across different languages – US/UK English, French, Finnish, Japanese and Mandarin – and the mechanisms underlying the substitutions.

Table 4.7: Proportion of variegated and harmonised child forms for variegated targets across various languages

Legend: VAR = variegation, CR = consonant repetition/CH = consonant harmony (use of CR denotes target form, use of CH denotes substitutions). Data from Finnish, French, US/UK English and Japanese are taken from Vihman et al. (2022). Data from Mandarin are from Lou (2020) and the present study (grey cells, previously presented in Table 4.5). Data are ordered in increasing order of the column '% VAR child forms for VAR targets'.

Child forms	N children	% VAR targets	% VAR child forms for VAR targets	% CR targets	% CH child forms for VAR targets
Finnish	5	0.57	0.08	0.15	0.27
French	5	0.57	0.27	0.10	0.20
US English	6	0.66	0.31	0.14	0.27
UK English	6	0.54	0.34	0.12	0.22
Japanese	7	0.42	0.50	0.05	0.21
Mandarin (Lou, 2020)	5	0.32	0.67	0.03	0.05
Mandarin (present study)	4	0.44	0.73	0.05	0.10
mean		*0.50*	*0.41*	*0.09*	*0.19*

Table 4.7 shows that the proportion of variegated *targets* in the Mandarin data from the present study was close to the mean, but the proportion of variegated *child forms* for variegated targets was the highest of all the languages (73%; with Lou's Mandarin data being the next highest at 67%). For consonant harmony, the opposite pattern can be observed: consonant repetition in targets in general is around 10% across all languages, but both Mandarin datasets show the lowest proportion of harmonised child forms for variegated targets (10% in present study; 5% in Lou, 2020). This pattern of prevalence of variegation for variegated targets and relatively little consonant harmony in Mandarin is interesting. In many studies of other languages, consonant harmony is discussed as a mechanism used by children *in response to* variegated targets (e.g., Menn, 1983). Yet, the Mandarin-learning children in the present study and in Lou (2020) are not using harmony as a mechanism; they are producing more variegated forms for variegated targets, relative to harmonised forms.

A closer inspection of the two Mandarin datasets revealed differences in the child responses to variegation that warranted further analyses. This is presented in Figure 4.4, but the differences in proportion of word structures in mothers' input in both Mandarin datasets are first presented in Table 4.8. Reference will be made to both table and figure following the figure.

Table 4.8: Proportion of word structures in mothers' input in Lou (2020) and the present study

Legend: RED = reduplication, VAR = variegation, CR = consonant repetition; proportions for the present study (grey cells) were previously presented in Table 4.5). **Bold** = referenced in text

mothers' input	N children	RED	CR	VAR	other	vocalic
Mandarin (Lou, 2020)	5	**0.15***		0.31	**0.08**	not reported
Mandarin (present study)	4	**0.09**	0.06	0.57	**0.26**	0.01

* Lou (2020) did not distinguish between RED and CR forms, so the RED percentage may have included some consonant-repeated forms.
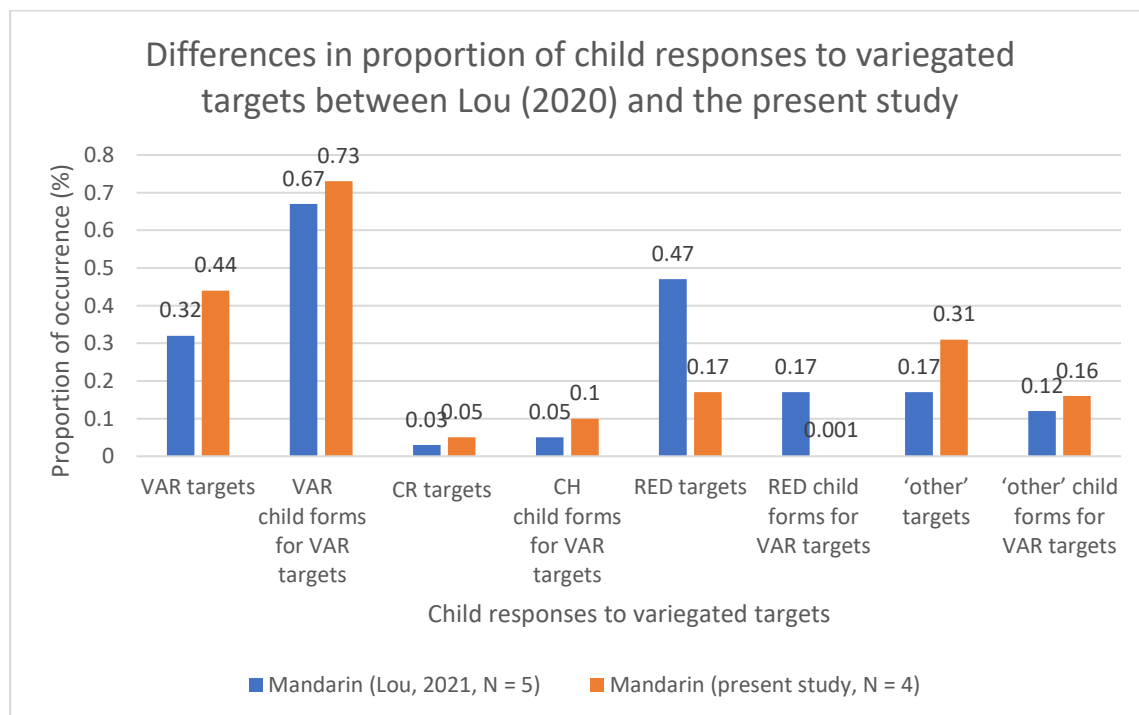
Differences in proportion of child responses to variegated targets between Lou (2020) and the present study

Legend: VAR = variegation, CR = consonant repetition/CH = consonant harmony (use of CR denotes target form, use of CH denotes substitutions), RED = reduplication. Data from present study was previously presented in Table 4.7.

Figure 4.4: Differences in proportion of child responses to variegated targets between Lou (2020) and the present study

Across the two Mandarin datasets, while the proportions of variegated responses for variegated targets are similarly high (first two sets of bars) and the proportions of harmonised responses for variegated targets are similarly low (third and fourth sets of bars), there is a difference between the proportions of reduplicated and 'other' targets and responses for variegated targets. The children in Lou (2020) are targeting almost three times as many reduplicated targets as the children in the present study, whereas the children in the present study are targeting almost twice as many 'other' targets as the children in Lou (2020) – seemingly a reverse pattern across these two types of word structures.

Several factors may account for the discrepancy between the two Mandarin datasets: differences in language environments, differences in the lexical stages of the children in the two datasets and cultural differences in the use of types of reduplications. First, Lou's child participants are heritage speakers learning Mandarin monolingually at home in Yorkshire, UK, where they are exposed to English exclusively in their ambient environment *outside* the home; in contrast, the children in the present study are immersed in a multilingual environment (mostly Mandarin and English, with some presence of additional Chinese languages) right from birth. This is because the bilingual education policy makes it compulsory for children to learn English and an official mother tongue in school (Stroud & Wee, 2010; see further discussion of language use in Singapore in section 4.2). Thus, even before their children are of school age, parents begin to speak to their children in both languages, to ensure that they will be on par with (or ahead of) other children by the time they do enter preschool. Thus, the Singapore children are also probably exposed to more English than Mandarin *inside* their homes. Hearing English in the home and ambient environment may thus have led to the differences in findings: recent work by Styles and colleagues (2022) have found that parents switch among languages about 5% of the time when speaking with their children at home. With regards to the lack of segmental templates found in the present study (since these forms were used merely as substitutions for sounds not yet acquired, precluding the definition of 'template'), perhaps there is a longer acquisition period for certain consonants as a result of being in a multilingual environment. Further work is warranted to tease apart the potential bilingual element in these findings: comprehensively computing the language mix of families and studying children at 2.5 to 3 years and beyond to examine whether consonant sequence templates may be found.

Second, although this study began recording the children 'at the 25wp', they are likely to actually be a bit beyond that point: some of them had already started combining words in the first session. A calculation of the mean length utterance of words (MLU) revealed that children were speaking between 1 and 2.5 words ($M = 1.7$) in their first recording session

and between range 1.5 and 3.5 words (*M* = 2.3) by the last recording session. Mothers may produce more reduplicated words when children are at an earlier lexical stage, consistent with corpora findings that reduplicative words were among the top 20% of the most frequent language input to children below 3 years old (e.g., Chang Corpus: Chang, 1998 and Tong Corpus: Deng & Yip, 2018 in the CHILDES database, MacWhinney, 2000). According to Ota and Skarabela (2016; 2018), reduplicated words are easier to learn (cf. Endress et al., 2007) because they embody a general rhythmic bias (cf. Thelen, 1981) for repeated elements (Vihman et al., 2023). Mothers of the younger and less lexically advanced children in Lou (2020) may thus have been producing more reduplicated (15%) than 'other' forms (8%, from Table 4.8), which may serve as a training ground (or what Caselli et al., 1995 term a "starter set", p. 19) from which children can learn more words. The mothers in the present study, whose children are more lexically advanced, have likely 'progressed' from using reduplicated (9%) to the 'other' word structure (26%, from Table 4.8), the proportions of which are thus reflected in the children's targets (and the children's own forms) (Figure 4.4). This discussion about the differences in the use of reduplication between the two Mandarin datasets continues in the third point below. Nonetheless, this study of children at a slightly later lexical stage provides a continuation and exploration of the kinds of patterns that children produce early in the word-combination period.

Third, the difference may be due to the language variety of Mandarin and cultural differences in the use of reduplications between Mandarin-speaking mothers from Mainland China and Singapore. Lin (1990) identified four types of reduplications, all of which appear in both Mandarin datasets (presented in Table 4.9, also revisited in section 4.2.3.2 under kinship terms). According to Lin (1990), reduplicated volitional verbs lead to attenuative forms (e.g., *tiŋ1* 'listen' → *tiŋ1tiŋ0* 'listen a little bit'), reduplicated descriptive adjectives lead to intensive forms (e.g., *hoŋ2* 'red' → *hoŋ2hoŋ2* 'very red'), reduplicated common nouns (specifically, numeral classifiers, Liu, 2012) lead to repetitive forms (e.g., *ɹən2* 'human' → *ɹən2ɹən2* 'everybody') and reduplicated kinship terms lead to vocative forms (e.g., *ma1* 'mother' → *ma1ma0* 'mother'). The last category of reduplicated terms presented in Table 4.9 – 'truncation then repetition' – is not in Lin (1990) but applies to Lou's (2020) data; it is a form of diminutive reduplication.

Table 4.9: Differences in categories of reduplicated words used in Lou (2020) and the present study

categories of reduplications (from Lin, 1990)	Lou (2020)	present study
	similar usage	
kinship/vocative	*pa4* 'father' → *pa4pa0* *ma1* 'mother' → *ma1ma0*	*pa4* 'father' → *pa3pa2**
adjective/intensive	*tʰoŋ4* 'painful' → tʰoŋ4tʰoŋ4 'a little painful' *kau1* 'high' → *kau1kau1* 'high'/indirect request to be carried up high	
	different usage	
verb/attenuative	*faŋ4* 'put' → *faŋ4faŋ0* 'put it [some place]' *ai4* 'love' → *ai4ai4* 'love love'	*pʰai1* 'tap' → *pʰai1pʰai1* 'tap it a little' *tsʰuei1* 'blow' → *tsʰuei1tsʰuei1* 'blow on it a little'
noun/repetitive	*fan4* 'rice' → '*fan4fan4*' 'rice rice *təŋ1* 'light' → *təŋ1təŋ1* 'light light'	*mau1* 'cat' → *mau1mau1*, similar to 'kitty' *kou3* 'dog' → *kou2kou3*, similar to 'doggy'
truncation then repetition	*ɕi2tsau3* 'shower' → *tsau3tsau0* 'shower'	none

* See section 4.2.3.2 for discussion of kinship terms

Based on Table 4.9, kinship terms and adjectival reduplications are common across both Mandarin datasets. For reduplicated verbs, mothers in the present study only used them for verbs that denoted a light touch (e.g., tap it a little, blow on it a little), they did not use it for verbs like 'love'. For reduplicated nouns, while other repetitive forms exist in the (adult) language (e.g., *ɹən2* 'person' → *ɹən2ɹən2* 'people', *tʰiɛn1* → *tʰiɛn1tʰiɛn1* 'everyday'), Lou's 'rice rice' and 'light light' examples are typical of child-directed speech; mothers in the present study did not reduplicate in the same way as those in Lou's study; they seemingly only used this form for words in the animal category (cf. *kitty*, *doggy*). Finally, none of the mothers in the present study produced the 'truncated, then repeated forms' which were diminutive reduplications; all produced such words in full. In short, children in Lou (2020) targeted four times as many types of reduplicated words as the children in the present study, probably because they are also receiving more reduplicated input from their mothers.

4.2 Tone analyses

Tone analyses were also conducted using the same three data sources mentioned for the segment analyses: what children hear, what they target and what they produce. Mothers' input speech provided a measure of the ambient language the children *heard*, targets refer to the adult forms that children were *trying to say*, e.g., *ta4ɕiaŋ4* 'elephant', and child forms refer to what children *produced*, e.g., *ta1taŋ1* for *ta4ɕiaŋ4* 'elephant'.

Specific to this section on tone analyses, the research questions addressed are:

1) Is there overlap among the tone sequences children hear, target and produce?
2) Is there evidence of templatic usage in the children's tone productions?

4.2.1 Tone sequences children hear, target and produce

A heat map table of the tone sequences children hear, target and produce is first presented in Table 4.10. This heat map will give us an idea of the frequency of occurrence of each tone sequence present in each of the three data sources. The main questions are to examine 1) if children's productions are similar to mothers' input and the words they target, as well as 2) if children are accurate in their tone productions, and if not, the kinds of substitute tone sequences the children make.

Table 4.10: Heat map table of proportions of occurrence of tone sequences in mothers'
input, child targets and child forms

Legend: S1 = first syllable, S2 = second syllable, % refers to the respective S1 or S2 sums over the
total number of tone sequences. Underlined text indicates highest proportion, darker shades indicate
larger proportions

mothers' input		S2					
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	S1 %
	Tone 1	0.07	0.04	0.03	0.07	0.04	0.26
	Tone 2	0.01	0.03	0.10	0.07	0.02	0.23
S1	Tone 3	0.07	0.07	0.00	0.09	0.02	0.26
	Tone 4	0.07	0.03	0.03	0.08	0.03	0.24
	Tone 0	0.00	0.00	0.00	0.00	0.00	0.01
	S2 %	0.24	0.17	0.17	0.32	0.10	

child targets		S2					
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	S1 %
	Tone 1	0.08	0.04	0.03	0.05	0.05	0.25
	Tone 2	0.03	0.03	0.12	0.07	0.02	0.28
S1	Tone 3	0.09	0.05	0.00	0.06	0.01	0.21
	Tone 4	0.05	0.02	0.03	0.08	0.07	0.25
	Tone 0	0.00	0	0	0	0	0.00
	S2 %	0.26	0.14	0.17	0.27	0.16	

child forms		S2					
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	S1 %
	Tone 1	0.14	0.04	0.06	0.14	0.17	0.55
	Tone 2	0.02	0.01	0.05	0.03	0.01	0.12
S1	Tone 3	0.07	0.08	0.01	0.04	0.00	0.20
	Tone 4	0.01	0.01	0.01	0.01	0.02	0.07
	Tone 0	0.04	0.00	0.00	0.01	0.01	0.06
	S2 %	0.28	0.15	0.12	0.23	0.21	

The findings from Table 4.10 are as follows:

1) Mothers produced T2-3 most frequently (10%), followed by T3-4 (9%) and T4-4 (8%). Mothers' input tended to have words with T3 or T1 equally in S1, although the four tones occurred fairly equally in S1 (with the exception of T0 which does not usually occur in S1 [see previous findings from Shang & Zhao, 2013; Zhu, 2002]). T4 tended to appear in S2 over the other tones.

2) Children targeted words with T2-3 most frequently (12%), followed by T3-1 (9%) and T1-1 and T4-4 (both 8%). Children tended to target words with T2 in S1, although, similar to mothers' input, the four tones also occurred fairly equally in S1. Again, similar to mothers' input, T4 in S2 slightly dominated over the other tones.

3) Children produced T1-0 most frequently (17%), followed by T1-1 and T1-4 (both 14%). More than half of the children's words were T1 in S1, and T1 tended to appear in S2 as well.

Thus, there is no common most frequent tone sequence among all three data sources of mothers' input, child targets and child forms. T2-3 is the common tone sequence between mothers' input and child targets and T1-1 is the common tone sequence between child targets and child forms. There is no common tone sequence between mothers' input and child forms (i.e., children are not producing the tone sequences they are hearing from their mothers). In general, the children produced T1-x tone sequence most frequently (making up 55% of all productions), especially the T1-0 tone sequence.

The set of contrasts in the tone sequences in Table 4.10 above is interesting as it shows that the tone sequences the children are hearing from their mothers and are targeting are not the tone sequences they are producing. The children may be targeting words with the same frequent tone sequence that they hear their mothers produce (T2-3), but the T2-3 tone sequence is not being produced frequently by the children. Meanwhile, children are targeting and producing the T1-1 tone sequence fairly frequently, potentially a *selected* tone template (cf. Mok & Lee, 2018; Vihman, 2019, see discussion in section 4.2.3.2).

Moreover, based on Table 4.10, while there were few instances of words with neutral tones, mothers and children in the longitudinal study did produce them, and the pitch of the neutral tone also differs depending on the tone of the preceding syllable, as has been previously reported. This is corroborated by acoustic analyses using Praat (Boersma & Weenik, 2012) – a sample of the T0 words produced by mothers (5a) and children (5b) is displayed in

Figure 4.5 below. (Not all the mothers and children produced T*x*-0 tone sequences; accordingly, some tone sequences are missing.)
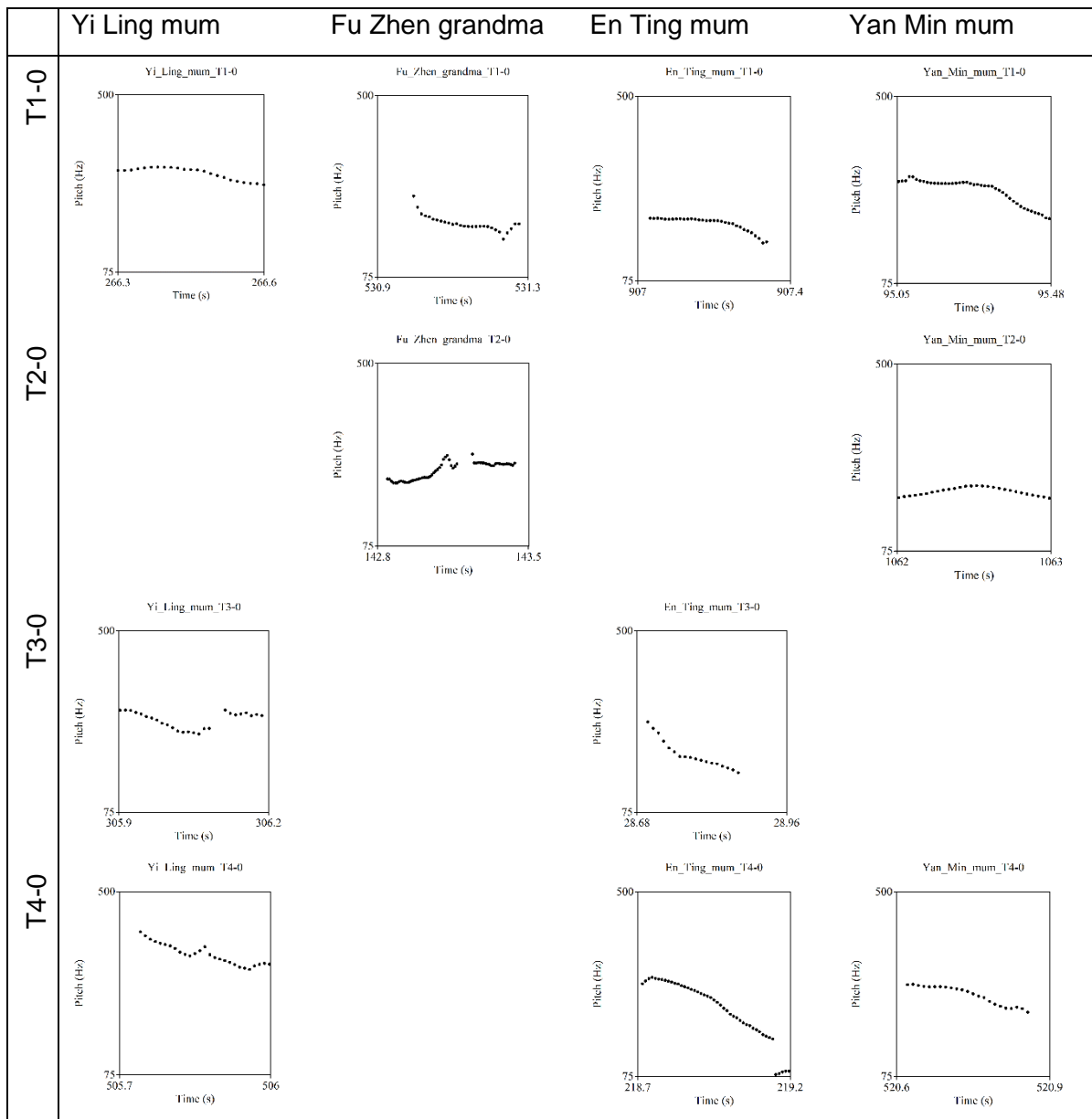
Figure 4.5a: T*x*-0 tone combinations produced by mothers

	Yi Ling	Fu Zhen	En Ting	Yan Min
T1-0				
T4-0				
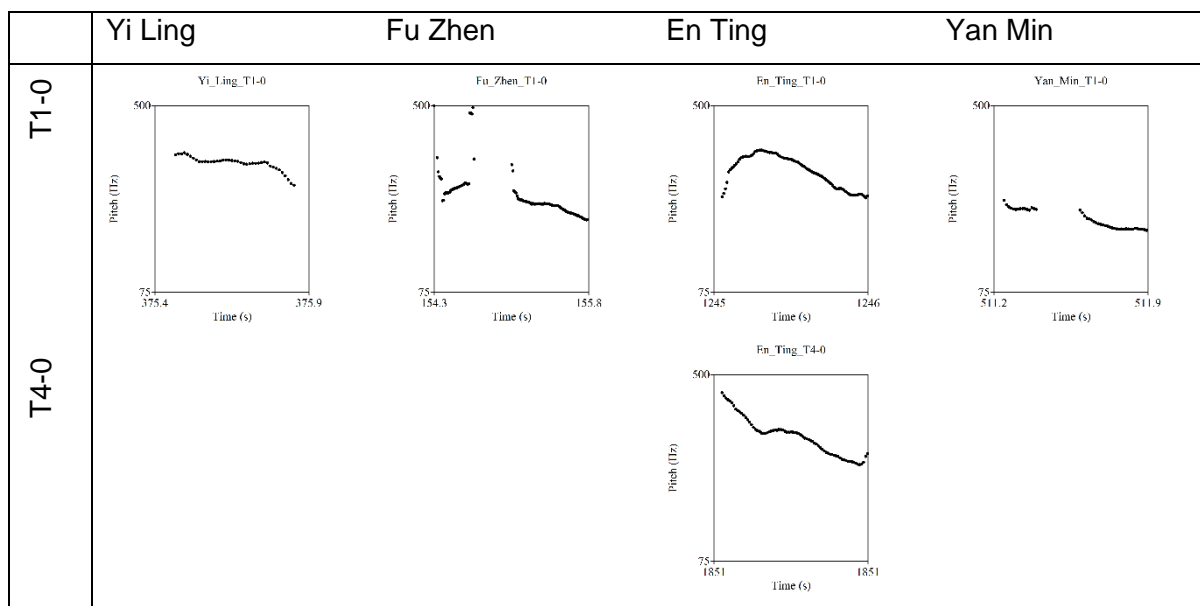
Figure 4.5b: T*x*-0 tone combinations produced by children

T1-0 tended to start high and level and then fall slightly, T2-0 began with a rise and then levelled off, T3-0 dipped before rising slightly and T4-0 started with a fall and continued to taper off. This finding of T0 in S2 in the present study contradicts previous discussion (Lock, 1989; Shang & Zhao, 2013, 2017), which concluded that there is no use of the neutral tone in Singapore Mandarin. Lock (1989) explained that Singapore Mandarin speakers may have retained more toned syllables due to the cross-linguistic influences of the syllable-timed manner of Southern Chinese languages native to many speakers in the Singapore Chinese community. Shang and Zhao (2013, 2017) examined the Mandarin education syllabus in Singapore and found no explicit pedagogical instruction in relation to neutral tone. They state that while students in classes may use the neutral tone when reading passages out loud, they do not typically produce it in everyday conversation. Additionally, the authors analysed the neutral tone usage of television broadcasters and found inconsistencies, e.g., the same word *ma1ma0* 'mother' was produced as *ma1ma1* in two instances and *ma1ma0* in three instances. The strong claim of 'no neutral tone in Singapore Mandarin' may need to be toned down: there DOES appear to be neutral tone usage in Singapore Mandarin, though it may be inconsistent.

4.2.2 Level of precision of tone sequences and patterns of tone substitutions

The next analysis examines the precision of the productions of tone sequences, i.e., whether the child forms matched the target tone sequences. Table 4.11 depicts a heat map of the accuracy distribution of tone sequences.

Table 4.11: Heat map of distribution of precision of tone sequences by children

Legend: S1 = first syllable, S2 = second syllable. Numbers in cells indicate proportion of accurate productions. Underlined text indicates highest proportion, darker shades indicate larger proportions

child forms		S2					S1 mean
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	
	Tone 1	0.57	0.48	0.35	0.62	0.27	0.46
	Tone 2	0.20	0.39	0.37	0.22	0.13	0.26
S1	Tone 3	0.44	0.63	NA	0.42	0.00	0.37
	Tone 4	0.17	0.29	0.12	0.14	0.36	0.22
	S2 mean	0.35	0.45	0.28	0.35	0.19	

From Table 4.11, in general, T1 was the most precisely produced in S1 (46%, as seen in the rightmost column 'S1 mean'), and T2 was the most precisely produced in S2 (45%, as seen in the bottom row 'S2 mean'). Precision for T4 in S1 was low (22%, rightmost column, bottom-most row), because the majority of occurrences of T4 in S1 were substituted as T1: of the total productions of T4-*x* tone sequence types, 72% were produced as T1-*x* tone sequences, with the next most frequent substitution being T4-*x* → T1-0. Considering tone sequences as a whole, using the above-chance level (5%) acquisition criteria adapted from Mok et al. (2020) for disyllabic tone sequences, children have actually acquired all tone sequences except T3-0, which is also the lowest frequently-occurring tone sequence based on the CDI (3%, previously presented in Table 3.4). However, the level of precision varies across tone sequences (e.g., T4-*x* tone sequences had the lowest precision levels because they were substituted with others).

Next, the discrepancies between the top three most frequent tone sequences (from Table 4.10) and the top three most precise tone sequences (from Table 4.11) in the child forms are discussed. T1-0, T1-4 and T1-1 were the most frequently occurring tone sequences, whereas T3-2, T1-4 and T1-1 were the most precisely produced tone sequences. The frequency × precision of these tone sequences is presented in Table 4.12.

Table 4.12: Frequency × precision of top tone sequences

tone sequence	frequency	precision	comment
T1-0	0.17	0.27	high frequency low precision
T1-4	0.14	0.62	high frequency high precision
T1-1	0.14	0.57	high frequency high precision
T3-2	0.08	0.63	low frequency high precision

Based on Table 4.12, T1-4 and T1-1 both occur frequently with high precision. This is unsurprising, given that both tone sequences were reported in Lou (2020) at an earlier lexical level, the 4wp and 25wp. The remaining two tone sequences are interesting: T1-0 has high frequency but low precision, because it was a substitute for many tone sequences. On the other hand, while T3-2 did not occur very frequently, it nonetheless had the highest precision, which could be because many of the T3-2 words were kinship terms. Children presumably could produce the terms of address of their daily interlocutors very precisely. In fact, T1-0, T1-1 and T3-2 were systematically produced among kinship terms; section 4.2.3.2 is dedicated to further discussion of the occurrence of these three tone sequences in kinship terms.

4.2.3 Discussion

The focus of this chapter was the tone sequences and systematic patterns or 'tone templates' in the children's speech. The findings from the present study provide an interesting perspective on the tone sequences the children target and produce: children's output did not mirror the input they received. The findings and the implications of these findings are explored in what follows.

4.2.3.1 High frequency and moderate level of precision of T1-x tone sequences in children's productions

In general, the tone sequences children produced were different from what they heard from their mothers and also different from words they targeted. Previous findings have been mixed regarding whether children between the ages of 1 and 3 have 'mastered'/acquired single tones (yes: e.g., Li & Thompson, 1977; Zhu, 2002; no: e.g., Wong & Strange, 2017; Xu Rattanasone et al., 2018). No study has looked at the mastery/acquisition of tone *sequences*: the findings of this study add on to the acquisition literature that children have acquired most tone sequences by age 2 years and up. The order of acquisition, according to Table 4.11, seems to be T1-*x* > T3-*x* > T2-*x* > T4-*x* sequences. Recall that Wong (2013) proposed the following order for acquisition of single tones: T4 > T1 > T2 > T3. It appears that T4 is first acquired as a single tone, but when it comes to sequences, T1-*x* sequences are more readily produced. The finding that T1-*x* sequences are more readily produced and with the highest level of precision is expected, because T1 is a well-practiced tone from the babbling years (cf. Lou et al., 2018) and earlier lexical points (cf. Lou, 2020). As a single tone, T3 has the most complex fall-rise tone contour of the four tones. In a tone sequence, especially when T3 is followed by T2, its tone contour is often reduced to the earlier dipping

2-1 part only (Li & Thompson, 1977; Zhu, 2002), thus the transition from the 2-1 in T3 to the 3-5 in T2 may be quite natural. T3-2 has the highest level of precision of all tone sequences, probably also because many of the terms with T3-2 are familiar kinship terms. This discussion on T3-2 and kinship terms continues in section 4.2.3.2.3.

Additionally, there was also a discrepancy between the frequency of occurrence of tone sequences and level of precision. There are varying levels of precision across the tone sequences, because many tone sequences were produced as substitutions for other tone sequences. This discrepancy is discussed in the next analysis: the findings of the frequency and precision of the most frequently produced tone sequences from the present study are compared with children at an earlier stage of lexical development – the 4wp and 25wp stages of Lou (2020), presented in Table 4.13 below.

Table 4.13: Frequency × precision of the most frequent tone sequences over three lexical stages

Legend: NA = not reported; table is ordered by the decreasing order of 'present >25wp frequency'. Data from 'Lou4wp' and 'Lou25wp' are from Lou (2020), data from the present study (grey cells) was previously presented in Table 4.12.

	Lou4wp		Lou25wp		present >25wp	
	frequency	precision	frequency	precision	frequency	precision
T1-0	0.20	0.55	0.18	0.41	0.17	0.27
T1-4	0.25	0.00	0.15	0.08	0.14	0.62
T1-1	0.13	0.57	0.14	0.23	0.14	0.57
T3-2	NA	NA	NA	NA	0.08	0.63

Across both studies (Lou, 2020 and the present study), T1-0, T1-1 and T1-4 were the most frequent tone sequences the children produced. However, the levels of precision differ. The level of precision of both T1-0 and T1-1 fell between Lou4wp and Lou25wp. Lou mentioned that the words produced at the 4wp consisted mainly of kinship terms and onomatopoeia (many of which are T1-0 and T1-1), which explained the high precision. At the 25wp, when kinship terms and onomatopoeia were still frequent, T1-0 and T1-1 were produced both for lexical targets and as substitutes for lexical targets with other sequences, which accounted for the reduction in precision.

At the 4wp, T1-4 was produced only as a substitute for other tone sequences, so at the 25wp there was some level of precision where previously there had been none, whereas precision for the T1-4 tone sequence was high in the present study. For the present study, T1-1 and T3-2 also showed high levels of precision, mostly because many of these words were kinship terms.

However, there are a few important differences between Lou (2020) and the present study, as mentioned in the discussion section earlier (section 4.1.4.2). In brief, Lou studied children learning Mandarin monolingually at home in Yorkshire, UK, while their ambient environment outside the homes had English exclusively, whereas the children in the present study were immersed in a multilingual environment (mostly Mandarin and English, with some presence of additional Chinese languages). Thus, some of these differences may be due to multilingual influences, especially in the case of kinship terms.

It is likely that the frequency and precision of tone production are also in part dependent on the physiological articulation of the tones themselves. Tone is a function of the rapid modulation of the vocal cord muscles, which tense or relax, depending on the tone (Ohala, 1978; Sagart et al., 1986). T1 is a high level tone whereas T4 is high falling ['55' and '51', respectively, in Chao (1930) notation]. T4 is articulated by decreasing vocal fold vibration or releasing the tension of laryngeal muscles, resulting in a natural falling contour. This could be why T4 occurs less often in S1; it is more common in S2 (see Table 4.11), the tone being a falling contour is more effortlessly produced in S2. This could also explain why the T4-*x* tone sequences in the present study were overwhelmingly produced as T1-*x* tone sequences, leading to low precision levels for T4-x tone sequences.

This set of findings follows from the differences Lou et al. (2018) found in tonal contours in the babbling of Mandarin- compared with English-learning 9- to 12-month-old infants (8 in each group). Mandarin-learning infants tended to use T1-4 and T1-1 and not T4-*x* tone sequences (similar to the overwhelming substitution of T4 in S1 with T1 found in the present study); Mandarin-learning children used significantly more T1 than the English-learning infants. In contrast, English-learning infants favoured T4-*x* tone sequences, especially T4-4, when they did produce disyllabic vocalisations (which were less often observed in that sample). Lou et al. concluded that the falling contour (T4) represents a 'global intonation' (given that English-learning infants also produce the tone), but to the Mandarin-learning infants, the tone also represents a lexical contrast.

Thus, Lou's (2020) finding that children in her study favoured tone sequences which start out high and then fall, i.e., T1-4 and T1-0, extends to the later lexical stage observed in the children in this study. Even though the present study did not test children from the 4wp, extrapolating from Lou (2020), the high frequency of use of both T1-*x* tone sequences found in the present study may have had its roots in the well-practiced motoric routines from these children's babble and early word production as well.

In terms of selection vs. adaptation (cf. Vihman & Croft, 2007; Vihman, 2019), the high frequency and moderately high levels of precision of T1-1 indicates that children *selected* T1-1 words to target and produce (as these are a frequently occurring tone sequence in both targets and child forms), then *adapted* other tone sequences to fit T1-1 (e.g., substituting T1-1 for T4-*x*). Just as Mok and Lee (2018) found for the templates in their study, the T1-1 tone sequence in this study was also used as an apparent template alongside the accurate adult form.

While T1-0 was the most frequent tone sequence in the child forms, it was also the least precise. T1-0 was seemingly the *de facto* substitute tone sequence, which might mean it was used as a tone template, similar to but different from the T1-4 Cantonese tone template Mok and Lee (2018) found in their study. The low frequency of occurrence of T1-0 in a child's ambient environment (indicated from mothers' input) and words the children target (see Table 4.13) shows that T1-0 was a tone template to which children *adapted* other tone sequences. Just as Choo et al. (2019) found and unlike Mok and Lee's (2018) original finding, the T1-0 tone template was not predominantly found in reduplicated forms. The reasons for the prevalence of the T1-0 tone sequence as a substitute tone sequence could thus be three-fold: not only is the sequence fairly effortless to produce in articulatory terms, the tone contour is also similar to the natural contour of English, the dominant language in the environment and it has been a well-practiced tone sequence since the babbling stage (cf. Lou et al., 2018).

4.2.3.2 Systematic patterns of tone production in kinship terms in Singapore Mandarin

The previous sections have revealed the prominence of the T1-1, T1-4 and T1-0 tone sequences in children's productions and the precision of T1-4 and T3-2 tone sequences. This is despite the fact that none of these were among the mothers' most frequently produced tone sequences, indicating a discrepancy between children's input and output. There was also a discrepancy between the tone sequences that were frequently produced as compared with those precisely produced by the children. T1-1 may be considered a tone template that was both selected and adapted, whereas T1-0 may be considered an adapted tone template. A closer look into the discrepancy reveals that three of these four tone sequences have a common lexical point of convergence: T1-1, T1-0 and T3-2 occur frequently in the category of kinship terms.

While kinship terms are usually disyllabic, they may stand alone as single syllables, termed the base syllable. Words in the familial category tend to be reduplications of the base syllable (e.g., in *ma1ma0* 'mother', the base syllable *ma* is repeated twice). As mentioned in section 2.2.2.1, reduplication is a productive morphological process (cf. Li & Thompson, 1981; Matthews & Yip, 2011). Lexical reduplication of both nouns and verbs has been found to be more common in Mandarin than in English (Xuan & Dolloghan, 2012). According to Lin (1990), there are four derivational types of reduplications in Mandarin (these have previously been mentioned in section 4.1.4.2) and her last category of reduplicated kinship terms is the current focus.

For mothers, kinship terms from other sessions were transcribed only if the terms had not been used in the first transcribed session. For children, kinship terms were collated across all sessions. Children's real names were omitted but nicknames were included.

The following tables show the mothers' (Table 4.14) and children's (Table 4.15) productions of kinship terms in this study.

Table 4.14: Mothers' productions of kinship terms

Legend: <u>Underlined text</u> indicates highest proportions

gloss	tone sequences										
	11	**14**	**10**	**20**	**31**	**32**	**41**	**42**	**44**	**40**	**01**
mother			*ma1ma0*			*ma3ma2*					
older brother						*kə3kə2* *kɒ3kɒ2*					
maternal grandfather	*koŋ1koŋ1*						*uai4koŋ1*[†]				
maternal grandmother						*pʰo3pʰo2*		*uai4pʰo2*[†]			
paternal grandfather				*ie2ie0*							
older sister					*tɕie3tɕie1* *tsɛ3tsɛ1*	*tɕie3tɕie2* *tsɛ3tsɛ2*					
paternal grandmother					*nai3nai1*	*nai3nai2*					
father			*pa1pa0*			*pa3pa2*				*pa4pa0*	
younger brother			*ti1ti0*							*ti4ti0*	
younger sister		*mei1mei4*	*mei1mei0*			*mei3mei2*					
child's nickname							*mi4kua1*				
child's bro's nickname			*tɕiŋ1kua0*								
child's nickname	*ɕi1ɕi1*		*ɕi1ɕi0*		*ɕi3ɕi1*	*ɕi3ɕi2*					*ɕi0ɕi1*
aunty						*a3i2*					
uncle									*tɕiu4tɕiu4*		
female maternal aunt						*piau3i2*					
N = 33 types	2	1	<u>6</u>	1	<u>4</u>	<u>12</u>	2	1	1	2	1
%	0.06	0.03	<u>0.18</u>	0.03	<u>0.12</u>	<u>0.36</u>	0.06	0.03	0.03	0.06	0.03

[†] See footnote of the use of *uai4* under section 4.2.3.2.1 'T1-1 pattern' below

Mothers mostly produced the kinship terms as T3-2 (36%), T1-0 (18%) and T3-1 (12%). However, there were certain terms that mothers produced in myriad different ways (e.g., there were four ways of producing of 'older sister' [*tɕie3tɕie1*, *tɕie3tɕie2*, *tsɛ3tsɛ1* and *tsɛ3tsɛ2*] and three ways of producing 'father' [*pa1pa0*, *pa3pa2* and *pa4pa0*]).

Table 4.15: Children's productions of kinship terms

Legend: grey cells – present in mothers' input (see Table 3.7) but not present in child forms, underlined text indicates highest proportions

gloss	11	12	14	10	20	31	32	34	41	42	40	01	02	04	`00
mother	ma1ma1		ma1ma4	ma1ma0	ma2ma0	ma3ma1	ma3ma2	ma3ma4				ma0ma1	ma0ma2		ma0ma0
older brother	kə1kə1 kɒ1kɒ1		kə1kə4 kɒ1kɒ4	kə1kə0 kɒ1kɒ0		kə3kə1	kə3kə2 kɒ3kɒ2					kɒ0kɒ1		ko0kɒ4	kʰɒ0kʰɒ0
maternal grandfather	koŋ1koŋ1			uai1koŋ0					uai4koŋ1						
maternal grandmother						po3po1	pʰo3pʰo2								
paternal grandfather															
older sister	tɛ1tɛ1 tɕe1tɕe1			tɕie1tɕie0		tɕie3tie1	tɕie3tɕie2 tsɛ3tsɛ2 te3te2					tse0te1			tə0te0
paternal grandmother						nai3nai1									
father	pa1pa1			pa1pa0			pa3pa2			pa4pa2	pa4pa0				
younger brother				ti1ti0 ti1tsi0 tsi1tsi0							ti4ti0				
younger sister															
child's nickname									mi4kua1						
child's bro's nickname				tɕiŋ1kua0											
child's nickname	ɕi1ɕi1			ɕi1ɕi0 i1ɕi0			tɕi3ɕi2 ɕi3ɕi2 i3ɕi2					hi0ɕi1			
aunty	a1i1	a1i2	a1i4	a1i0			a3i2								
uncle				diu1dou0											
female maternal cousin							pia3i2								
N = 62 types	<u>9</u>	1	4	<u>14</u>	1	5	<u>13</u>	1	2	1	2	4	1	1	3
%	<u>0.15</u>	0.02	0.06	<u>0.23</u>	0.02	0.08	<u>0.21</u>	0.02	0.03	0.02	0.03	0.06	0.02	0.02	0.05

The different ways children produced the kinship terms were mostly T1-0 (23%), T3-2 (21%), and T1-1 (15%). The remaining tone sequences each saw under 10% occurrence per type.

Taking Tables 4.14 and 4.15 together, T3-2 was the most frequently produced tone sequence by mothers and children in terms of highest total combined percentage. Children (23%) produced T1-0 proportionally more often than the mothers did (18%) as an adapted tone template, as discussed above; in the category of kinship terms, T1-0 seems to be an adapted tone template as well.

T3-1 was mothers' next most frequent tone sequence (12%), while T1-1 was the children's next most frequently produced sequence (15%), a sequence that was produced by mothers in only two words (*koŋ1koŋ1* 'maternal grandfather' and *ɕi1ɕi1*, Yan Min's nickname). T1-1, then, seems to be both a selected and adapted tone template in general and in the category of kinship terms as well.

These three tone sequences were also the only ones with more than 10% frequency in the children's forms. All other tone sequences were used less than 6% of the time. Subsequent analyses will focus on T1-1, T1-0 and T3-2, discussed in increasing order (by total percentage across mothers and children) of overall use.

Table 4.16 presents the kinship terms that were produced in the three tone sequence patterns.

Table 4.16: Observable patterns in children's productions of kinship terms

gloss	mothers' input	patterns in child forms		
		T1-1	T1-0	T3-2
mother	ma1ma0 ma3ma2	ma1ma1	ma1ma0	ma3ma2
father	pa1pa0 pa3pa2 pa4pa0	pa1pa1	pa1pa0	pa3pa2
older sister	tɕie3tɕie1 tsɛ3tsɛ1 tɕie3tɕie2 tsɛ3tsɛ2	tɛ1tɛ1 tɕe1tɕe1	tɕie1tɕie0	tɕie3tɕie2 tsɛ3tsɛ2 te3te2
older brother	kə3kə2 kɒ3kɒ2	kə1kə1 kɒ1kɒ1	kə1kə0 kɒ1kɒ0	kə3kə2 kɒ3kɒ2
Yan Min's nickname	ɕi1ɕi1 ɕi1ɕi0 ɕi3ɕi1 ɕi3ɕi2	ɕi1ɕi1	ɕi1ɕi0 i1ɕi0	tɕi3ɕi2 ɕi3ɕi2 i3ɕi2
maternal grandfather	koŋ1koŋ1 uai4koŋ1	koŋ1koŋ1	uai1koŋ0	
aunty	a3i2	a1i1		a3i2
younger brother	ti1ti0 ti4ti0		ti1ti0 ti1tsi0 tsi1tsi0	
Yi Ling's brother's nickname	tɕiŋ1kua0		tɕiŋ1kua0	
uncle	tɕiu4tɕiu4		diu1dou0	
maternal grandmother	pʰo3pʰo2 uai4pʰo2			pʰo3pʰo2

Two findings of note can be observed from this table. First, mothers themselves produced kinship terms using various tone sequences, so they too were inconsistent in their tone sequences. Children then sometimes *selected* the same tone sequences to produce, and at other times *adapted* other tone sequences to the ones already prevalent in their repertoire.

Next, different words lend themselves to different tone sequences: frequent terms have been known to be the most variably produced (cf. Stoel-Gammon, 2011), in this study, by children AND mothers alike. 'Mother', 'father', 'older sister', 'older brother' and Yan Min's nickname were the most frequently occurring kinship terms, which could explain why children produced them in all three observable patterns of tone sequences. These three tone sequences were both accurately produced and used as substitutions for the other two tone sequences by children. In addition, the presence of a referent in the situational context may also have affected the frequency of productions. Of these three words, 'older sister' was produced by all four children because of the investigator's presence: The parents had told their children to address the investigator as such (she was also referred to as 'aunty'[6] by Fu Zhen's grandmother and Yan Min's mother; interestingly Yan Min also referred to the investigator as 'older sister' at various times, without prompting from her mother). In general, the kinship terms used were mostly those of the immediate interlocutors.

4.2.3.2.1   T1-1 tone template

In general, from the earlier tables, T1-1 did not appear among mothers' frequent tone sequences, but was among the frequent tone sequences children targeted and produced (see Table 4.10). Specifically for kinship terms, there were only two T1-1 words mothers produced: *koŋ1koŋ1*[7] 'maternal grandfather' and *ɕi1ɕi1*, Yan Min's nickname. Thus, within the seven words with the T1-1 pattern in Table 3.11, five were produced with T1, either fully

---

[6] 'Aunty' is a term of politeness used in Singapore (and in many other places) to address middle-aged/matronly women "who are perceived to be old enough to be one's mother" (Wong, 2006, p.3). It is more often used for non-biological relations as biological aunts have specific titles depending on family conventions. It may also be used for younger women (e.g., in the case of Yan Min addressing the investigator), as parents do not want to confuse children who do not yet know the difference between biological 'older sisters' and 'older sisters' who may look too "youthful" to be an 'Aunty'.

[7] Maternal grandparents are addressed as *uai4pʰo2* 'maternal grandmother' (not produced in present study) and *uai4koŋ1* 'maternal grandfather'; paternal ones are *nai3nai1* 'paternal grandmother' and *ie2ie0* 'paternal grandfather' (only produced by mothers, not produced by children in present study). The prefix *uai4* means 'outside/external', reinforcing the patriarchal notion that the paternal line is 'inside/within' the family whereas maternal relations are considered 'outside/external [to]' the family. Some families choose to make it more neutral or to match the paternal side by simply reduplicating the *pʰo2* or *koŋ1* (see examples in Tables 4.14 and 4.15) or they may use something completely different (e.g., Fu Zhen calls her paternal grandmother *nai3nai1* but maternal grandmother *ma2ma3* [tone numbers as approximated to Mandarin]).

reduplicated ('maternal grandfather' and Yan Min's nickname), in either the first ('mother' and 'father') or the second syllable ('older sister').

Furthermore, the remaining two words, 'older brother' and 'aunty', can also include T1: the base syllable of 'older brother' is *kə1;* the reduplicated form can be *kə1kə1* (although mothers produced the T3-2 tone sequence more often and the *a* in *a3i2* 'aunty' is also often produced as *a1*. Even though the mothers did not produce these with T1 forms, the children could have heard these words with T1 forms from others, because the children did produce 'aunty' with *a1* as well (see Table 4.15). Therefore, the strong prevalence of T1 in these words may have led to the whole category of 'family members with whom the children come into daily contact' to follow a T1-1 tone sequence.

In short, T1-1 seems to be a unique tone sequence that children selected for production and adapted for non-T1-1 tone sequences. T1 is acoustically salient (Liu et al., 2007; Mok & Lee, 2018) and a tone that children are said to acquire first (Li & Thompson, 1977; Lou, 2020; Zhu, 2002). As mentioned earlier, T1-1 may have been a well-practiced tone sequence since the babbling stage for the children in this study as well (cf. Lou et al., 2018), so children are adept at producing T1 and, in fact, *over*producing them (see high precision of T1-1 even at the 4wp in Lou's data, presented in Table 4.13). The high frequency of occurrence and use of T1-1, mostly with reference to the children's immediate kin, may have led to its high precision, which in turn led to its becoming a tone sequence that children selected and adapted for use.

4.2.3.2.2   T1-0 tone template

Four of the items in the T1-0 set were produced as T1-0 by both mothers and children ('mother', 'father', 'younger brother' and Yan Min's nickname). These words relate to immediate family members with whom the children come into daily contact. Even though the children in the study did not produce 'father' frequently, the terms 'mother' and 'father' frequently occur together. The fact that 'mother' was produced more than twice as often as 'father' may have led 'father' to be adapted to having T1 in S1. The T1-0 tone sequence may then have generalised to words in the semantic category of 'family members', including all the other words in this set. In the specific category of kinship terms, T1-0 is thus a tone template that children both selected ('mother, 'younger brother', Yan Min's nickname) and adapted ('father').

Among the four words produced by mothers and children as T1-0, Yan Min's nickname and 'mother' have T1 base syllables whereas 'father' and 'younger brother' have T4 base syllables. As previously mentioned, T4 is frequently substituted with T1 in S1, which could also explain why 'father' and 'younger brother' were produced by both mothers and children with the T1-0 tone sequence. The T4→T1 substitution pattern in combination with the cross-linguistic transfer of the trochaic English pattern previously mentioned (cf. Mok & Lee, 2018) could then perhaps explain the child productions of *uai4koŋ1* 'maternal grandfather' and *tɕiu4tɕiu4* 'uncle' as *uai1koŋ0* and *diu1dou0,* respectively, again, an adaptation of T1-0 to non-T1-0 terms*.*

Additionally, Yi Ling's mother produced the most T1-0 tone sequence of all adults, which would explain why Yi Ling herself also produced the most T1-0 of all children. T1-0 made up 41% of her utterances, most of which referred to her immediate family members – 'mother', 'father' and 'younger brother'. Hence Yi Ling was matching what she heard in her input.

As mentioned earlier, T1-0, like T1-1, may also have been a well-practiced tone sequence since the babbling stage for the children in this study (cf. Lou et al., 2018).

4.2.3.2.3   T3-2 pattern

The T3-2 pattern found amongst the kinship terms is likely the most interesting of all. T3-2 is best considered simply a 'pattern' rather than a template, because it is an entrenched pattern of pronunciation in parents' usage as well and not an idiosyncratic child invention, but seemingly restricted to kinship terms and specific to Singapore, with its large amount of language mixing.

On the one hand, there may simply be a phonetic basis for the frequency of use of this pattern. The tone contour of T3 is a mild dip followed by a rise, described as 2-1-4 in Chao (1930) notation, while the tone contour of T2 is a rise from 3 to 5. When T3 is followed by T2, its tone contour is often reduced to the earlier dipping 2-1 part only (Li & Thompson, 1977; Zhu, 2002), thus the transition from the 2-1 in T3 to the 3-5 in T2 may be quite natural. As mentioned earlier, Wong (2013) also proposed that T3-2 has a simple F0 contour because there is less F0 transition from the low F0 offset in T3 to the low F0 onset in T2. This may be seen in the T3-2 acoustic contours produced by the mothers and children in Figure 4.6, where, in general, there is a small dip followed by a small rise.
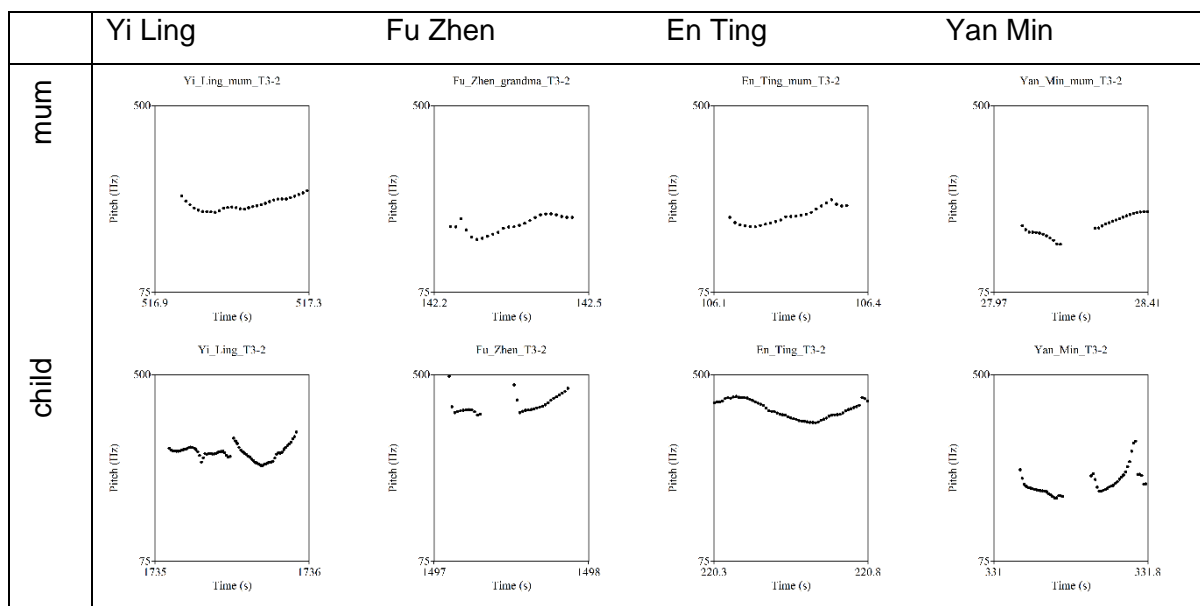
Figure 4.6: Acoustic contours of T3-2 produced by mothers (top) and children (bottom)

As mentioned under Figure 2.1, Lee (2010) ran an acoustic analysis of Singapore Mandarin and found that for T2, there was a fall to a mid-low stretch before a rise to a high pitch level at the end of the syllable. This pattern can be seen in some of the contours of two of the children, Yi Ling and Yan Min.

On the other hand, aside from articulatory properties, there may be a sociolinguistic explanation for the prevalence of the T3-2 pattern. The ethnic Chinese community in Singapore is a linguistically diverse one; there is a lot of language mixing (i.e., people use words from multiple languages in the same sentence) of the Chinese languages (even before adding English into the mix) as people from the different Chinese language groups inter-marry (and there are also, increasingly, inter-ethnic marriages, but this is beyond the scope of this thesis) and still maintain the languages of their homes in their new families. Matthews and Hickey (2010) discuss the influence of Mandarin syntax on the other Chinese languages, such as Cantonese and Hokkien (it can be argued that there is a corresponding influence of Cantonese [cf. Zhang, 1988] and Hokkien lexis [cf. Chen, 1983] on Mandarin as well [see also Chua, 2003; Lim, 2007]). Here the investigator speculates, similarly, that the T3-2 pattern of substitutions is a crossover effect from *Cantonese to Mandarin kinship terms* in Singapore.

The ethnic Chinese community in Singapore are made up mostly of people who came from various regions of southern China: the Hokkiens (39.6%), Teochews (19.8%), Cantonese (14.4%), Hakkas (8.4%), Hainanese (7.2%), Foochows (1.9%), Henghua (0.9%),

Shanghainese (0.7%), Hockchia (0.6%) and others (7.1%) (Department of Statistics Singapore, 2020). At independence (in 1965), no ethnic Chinese used Mandarin as their predominant home language (Afrendas & Kuo, 1980). For many Chinese families in Singapore, one's elders in the family are still addressed in non-Mandarin terms. According to Lee (2011), the use of these non-Mandarin terms is the result of the parents (e.g., the parents of the children in this study) having grown up with those languages in the household. The linguistically heterogeneous Chinese community thus spoke mutually unintelligible languages, which reinforced their segregation into their respective clan associations (Stroud & Wee, 2007), thus the Speak Mandarin Campaign was launched in 1979 to unite the community.

Nonetheless, some families still prefer to use non-Mandarin address terms as it is the language of the home and brings with it a sense of emotional belonging that runs deeper than would a Mandarin term of address (Lee, 2011), the latter being seen as a government-mandated language for which the older generation had to 'sacrifice' their own language (Rappa & Wee, 2007; see also Ng, 2010, for further discussion and a critique about the government policing what people could do in their personal households). Titles are used rather than names to address elders as a sign of respect, unlike the use of direct names in the western world. Even if names are used, a suffix that denotes one's relation to the other person is attached to it, e.g., *Yvonne piau2tɕie3* 'older maternal female cousin Yvonne'. If parents come from one of the different language groups (e.g., Hokkien, Teochew, Cantonese etc.), the titles used may reflect the language of their respective households. Thus, children may learn to address their Cantonese grandmother by the Cantonese term *ma:4ma:4* (tone numbers in Cantonese) and their Hokkien grandmother by *'a-ma'*. The high frequency with which children have to address family members of differing language groups and the corresponding intermingling of the Chinese languages might have thus created crossover patterns out of the various tone sequences in the respective languages.

Even though the Cantonese-speaking population did not (and still do not) make up the majority of the Chinese population in Singapore, Cantonese had a substantial influence on Singapore English in the 1980s (and its colloquial form Singapore Colloquial English, especially with regards to the use of particles, see Ansaldo, 2010 and Lim, 2007). Although Lim (2007) focuses on Singapore English, the language environment Lim is analysing is the same environment as in the present study. Lim (2007) proposed that the 1980s was the age of the 'global-media languages', including English and Cantonese (prior 'ages' included the influence of Malay, Hokkien and Tamil among others). Lim noted that Cantonese was particularly prominent for two reasons. First, culturally, the 1980s were the golden years of

music and film entertainment from Hong Kong, where martial arts films, television series and Cantonese pop music permeated the airwaves of Hong Kong and overseas Chinese communities (Wong, James, 2003, cited in Lim, 2007; see also Zhang, 1988).

Next, on a more linguistically relevant note, Cantonese continued to be used at home despite the Speak Mandarin Campaign (which had made most others switch to Mandarin as the home language over the other Chinese languages). When each language group was studied separately (most recently, in Lee, 2001), the Cantonese still spoke Cantonese more frequently at home (36% in 2000) compared with Mandarin (32%), in contrast to the Hokkiens and Teochews, who spoke more Mandarin (46% and 43%, respectively) than their own language (29% and 26%, respectively). Perhaps the influence of Cantonese can be observed from the fact that parents have generally begun to teach children the Cantonese term for 'older brother' and 'older sister' (regardless of the actual language group of the family – this applies to the families in this study as well as the general Chinese community in Singapore), while other terms of address are in Mandarin or the other Chinese languages.

For the T3-2 kinship terms, there was a full match between mothers' and children's productions: none of the T3-2 tone sequences children produced were a substitution of another pattern; they were all instances their mothers had also used. To better explain the T3-2 pattern in Mandarin, the parallels between the tones in Cantonese and Mandarin are presented in Table 4.17.

Table 4.17: Parallels between the 6 tones in Cantonese (adapted from the table in Yu, 2009) and 4 tones in Mandarin; Chao (1930) notations are used

| **Cantonese** | T1 53 ~ 55 | T2 25 | T3 33 | T4 21 | T5 23 | T6 22 | |
| **Mandarin** | T1 55 | T2 35 | | T3 214 | | | T4 51 |

This analysis utilises a similar way of discussing Mandarin vocative reduplications (cf. usage of the term by Lin, 1990) as the discussion of the same in Cantonese in Yu (2009). From the table, some similarities in the tone properties of Cantonese in relation to Mandarin may be observed: T1 ≈ T1, T2 ≈ T2 and T4 ≈ T3. Given that three of four Mandarin tones overlap with Cantonese, it is reasonable to expect an influence of Cantonese on some words in Mandarin. Table 4.18 lists the kinship terms that were produced as T3-2 in Mandarin and their equivalents in Cantonese.

Table 4.18: Comparisons between Cantonese[8] and Mandarin productions of kinship terms

Legend: grey cells: present in mothers' input (see Table 4.14) but not present in child forms

similar tone sequences (T4-2 ≈ T3-2)			different tone sequences (T4-1 → T3-2)		
gloss	Cantonese (T4-2)	Mandarin (T3-2)	gloss	Cantonese (T4-1)	Mandarin (T3-2)
(neutral) grandmother	*pʰo4pʰo2*	*pʰo3pʰo2*	mother	*ma:4ma:1*	*ma3ma2*
paternal grandfather	*ie4ie2*	*ie3ie2*	father	*pa:4pa:1*	*pa3pa2*
younger sister	*mui4mui2*	*mei3mei2*	older sister	*tsɛ:4tsɛ:1*	*tsɛ3tsɛ2*
younger brother‡	*tai4tai2* *ti4ti2*	*ti3ti2*	older brother	*kɔ:4kɔ:1*	*kɔ3kɔ2*

‡ Younger brother is a special case: *sai3lou2* is used more often than *tai4tai2*; *ti4ti2* may sometimes be used in child- or infant-directed speech. Separately, *ti4ti2* is sometimes used to refer to young boys in general, though the more common form is *sai3lou6*.

The frequency of the tone sequences T3-2 vs. T3-1 (which approximates the Cantonese T4-1) in mothers' speech also cannot explain why T3-2 is more prominent: the mothers make only 7% use of T3-1 and T3-2 (see Table 4.10).

Given the ubiquity of families teaching their children the Cantonese terms for 'older brother' and 'older sister', these words are the likely sources of the T3-2 tone sequence. However, this tone sequence is not exactly the same as in Cantonese, because in Cantonese the second syllable has a high tone, **T1**, i.e., *tsɛ:4tsɛ:1* and *kɔ:4kɔ:1* (see Table 4.18), whereas the Mandarin form uses the rising **T2**. Thus, the final Mandarin iteration that mothers produce and children learn in Singapore is a combination of *Cantonese tsɛ/kɔ syllables* with a *Mandarin T3-2 tone sequence.* However, the approximated Mandarin T3-2 tone sequence can nonetheless be found in other words in Cantonese (left block of Table 4.18), so it could have been adapted from those words and does not constitute a unique Mandarin creation: 'grandmother', 'paternal grandfather', 'younger sister' and 'younger brother' all have the Cantonese T4-2 tone sequence that is similar to the Mandarin T3-2 tone sequence. This Mandarin T3-2 tone sequence is then also adapted to the other two Cantonese T4-1 terms for family members to form *ma3ma2* 'mother' and *pa3pa2* 'father' in Mandarin.

---

[8] I thank Justin Lo for his assistance with the Cantonese forms.

Relating back with the T1-1 pattern, the fact that *tsɛ4tsɛ1* 'older sister' in Cantonese is similar to a T3-1 tone sequence in Mandarin (T4 in Cantonese is similar to T3 in Mandarin, recall Table 4.17) could also explain the T1-1 reduplication of the term. Children first picked up on the more salient T1 in the *tsɛ3tsɛ1* and *tɕie3tɕie1* 'older sister' productions by mothers; T1 is also in the second syllable position which makes it acoustically salient and recent, then reduplicated the T1 to produce *tɛ1tɛ1* and *tɕe1tɕe1* for 'older sister'. Given the closeness of semantic relations between 'older sister' and 'older brother', the T1-1 for 'older sister' could also have been adapted from *kə1kə1* and *kɒ1kɒ1* for 'older brother' ('older brother' itself also has a base syllable of *kə1*, as mentioned in section 4.2.3.2.1).

Referring to the T3-2 patterns in Table 4.16, Yan Min's mother also adapted the T3-2 tone sequence to her nickname *ɕi3ɕi2*. An extension to this T3-2 tone sequence is that all the adults except En Ting's mother adapted this tone sequence *back* to the Mandarin syllables for 'older brother' and 'older sister': the adults produced them as *kə3kə2* and *tɕie3tɕie2* and so did the children.

Note that this T3-2 Mandarin tone sequence has even been extended into English. Wong, Jock (2003) showed that terms like *boy3boy2* and *girl3girl2* (tone numbers in Mandarin) are also used for referring to little boys and girls in general in Singapore English, and may have been directly translated from the corresponding terms in Cantonese: *tsai4tsai2* 'boy' and *nœi4nœi2* 'girl' (tone numbers in Cantonese). In the present study the following English words were also produced by mothers in the T3-2 Mandarin tone sequence: *baby, ball ball, bear bear, darling, dog dog* and *mummy*. Woon and Styles (2021) have documented a dictionary of words used in routines with children in Singapore, some of which were produced by the mothers in this study in the T3-2 Mandarin tone sequence as well: *shee shee* 'pee', *ng ng* 'poo', *pom pom* 'shower', among others.

In Lou (2020) the mothers did not often produce T3 or T3-2, nor did the children in either the 4wp or the 25wp, likely due to the more challenging articulatory contour of T3 and also the less linguistically diverse environment (mostly monolingual Mandarin at home with more English outside the home).

Lou (2020) found monosyllabic tone templates (T1 and T4) in her children's productions at the 4wp but not the 25wp and no evidence of disyllabic tone templates at both word points. In the present study, monosyllables were not studied but there *was* evidence of disyllabic tonal templatic use by these children beyond the 25wp. As mentioned at the end of Chapter 2 as well, this discussion will be left to the final chapter in order to track the developmental

timeline and rise and fall of templatic use from the younger and less lexically advanced children in Lou (2020) through to the older and more lexically advanced children in the present study.

In the next chapter, the same segment and tone analyses as have been discussed in this chapter will be examined in an experimental study of Mandarin nonword repetition. Additionally, the possible influences of the frequency of segmental sequences and tone sequences on children's productions of segments and tones in newly encountered words are studied, followed by an exploration of what these findings tell us about phonological development in Mandarin.

Chapter 5 – Experimental study: Results and Discussion

5.1 Segmental analyses

The previous chapter revealed systematic patterns of segment and tone productions and substitutions in children's everyday speech. In order to provide in-depth insight into children's responses to words they have never encountered before, the same segment and tone analyses that were carried out on the naturalistic data were also carried out with the nonwords of a nonword repetition task. The similarities and differences in patterns and substitutions of segmental sequences and tone sequences between the two paradigms can then be compared; this comparison will be addressed in Chapter 6.

5.1.1 Manner categories of segmental sequences in disyllabic nonwords

The focus of this section is on the segmental sequences of the disyllabic nonwords. Table 5.1 provides a summary of the most common responses and substitutions in terms of the frequency of manner categories of segmental sequences (Table 5.1).

Table 5.1: Most frequent manner categories in child responses to nonword targets

Legend: P = plosive, F = fricative, N = nasal, L = liquid, V = vowel. Percentage (%) refers to the proportion of all child responses produced as the respective manner categories. Grey cells indicate infrequent segmental sequences based on CDI

nonword targets (*N* = 32, 4 per category)	child responses			
segmental sequence	most frequent	%	next most frequent	%
NV-PV	NV-PV	0.52	V-PV	0.26
LV-FV	V-FV	0.43	LV-FV	0.19
PV-PVN	PV-PVN	0.42	PV-PV	0.16
FV-FVN	FV-FVN	0.39	FV-FV	0.16
FV-NV	FV-NV	0.47	PV-NV	0.23
PV-NV	PV-NV	0.31	PV-V	0.17
PV-FVN	PV-FVN	0.41	PV-FV	0.20
LV-PVN	V-PVN	0.40	LV-PVN	0.17

* Note: Other manner categories children produced were under 10% so they were not included in this table.

The average level of precision for segmental sequences was about 40%. Of the 32 nonword targets, 12 (37.5%) were plosive-initial, 8 (25%) each were fricative- and liquid-initial and 4 (12.5%) were nasal-initial. This corresponded roughly to the distribution as the children experience in the language, based on the Singapore CDI. Overall, plosive-initial responses occurred the most often, followed by zero-onset/vowel-initial, fricative-, nasal- then liquid-initial responses. The fact that plosive-initial responses were produced the most frequently corresponds to the relative frequency of these nonword targets in the task. However, this high frequency is also due to the fact that plosives often substitute for fricative-initial nonword targets. This stopping process has also been observed in other analyses of children's words in Mandarin (cf. Cho, 2008; Jeng, 2011, cited in Li & To, 2017).

5.1.2 Children's responses to disyllabic nonwords with various word structures

The word structures of the children's responses to the nonwords were explored next. The findings from the previous section showed that children were fairly accurate in their production of the word-initial consonant, depending on the consonantal manner categories of the nonword targets, and the substitution pattern seemed to be mostly omission of the word-initial consonant. The following analyses of the word structures of the children's responses to the nonwords were also carried out so that the findings can be compared with the data from the naturalistic observations (see Chapter 4).

Of the 32 nonword targets, 24 (75%) were variegated and 8 (25%) were 'other'. Table 5.2 provides a summary of the distribution in word structures of the nonwords and the children's responses. In the data presentation, the categories of reduplication and consonant harmony took precedence over the rest, i.e., if there were 'other' or 'vocalic' words that were reduplicated or harmonised, they were grouped in the reduplicated or harmonised categories.

Table 5.2: Word structures of disyllabic nonwords and child forms, classified by segmental sequence

Legend: P = plosive, F = fricative, N = nasal, L = liquid, V = vowel; RED = reduplication, CH = consonant harmony, VAR = variegation, voc = vocalic, 1 syll = number of responses that were one syllable in response to the disyllabic nonword. Underlined text (bottom row) indicates highest proportion in child forms. **Bold**/**bold and underlined text**: referenced in text

		nonword targets	**child forms**					
	n	segmental sequence	RED	CH	VAR	other	voc	1 syll
VAR sequences (*N* = 24)	4	PV-NV	0	0.02	0.61	0.33	0	0.04
	2	PV-FVN	0	**0.20**	0.60	0.20	0	0.00
	4	PV-PVN	0	0.13	0.71	0.13	0	0.03
	1	FV-FVN	0	0.07	**0.33**	**0.47**	0	0.13
	3	FV-NV	0	0.00	0.63	0.27	0.02	0.08
	4	NV-PV	0	0.05	0.62	0.32	0	0.02
	2	LV-FV	0	0.03	**0.32**	**0.61**	0	0.03
	4	LV-PVN	0	0.03	**0.33**	**0.60**	0	0.03
'other' sequences (*N* = 8)	2	PV-FVN	0	0.07	0.14	0.76	0.03	0.00
	3	FV-FVN	0	**0.26**	0.12	0.60	0	0.02
	1	FV-NV	0	0	0.13	0.75	0.06	0.06
	2	LV-FV	0	0.03	0.16	0.78	0	0.03
		total	0	0.07	0.45	0.44	0.01	0.03

Note: the fricative in the 'other' sequences is /h/, a glottal fricative, hence it is considered an 'other' sequence (made up of only one supraglottal consonant, which is the P or F or L in the sequences)

For the variegated sequences, most were produced as variegated except FV-FVN, LV-FV and LV-PVN sequences (**bold** text): these were produced as 'other'. For all three sequences, the word-initial consonant was omitted, i.e., the first fricative in FV-FVN and the liquid in the other two sequence. For the 'other' sequences, most were produced as 'other'.

There was some harmony found for the nonword targets as well. All the harmonised forms were substitutions because there were no consonant repetitions amongst the nonword targets. The largest proportion of harmonised responses were in response to the variegated PV-FVN sequence and the 'other' FV-FVN sequence (**bold and underlined** text). For the variegated PV-FVN sequences, the harmonised responses were [t]-[t] forms; for the 'other' FV-FVN sequences, the harmonised responses were a range of fricative consonants.

As in the naturalistic findings (Chapter 4), there were no reduplicated responses, and the responses were almost equally split between variegated and 'other' responses (45% and 44% respectively). This is unsurprising, given that the two categories of nonwords fell into just these two categories. Additionally, the analyses in Chapter 4 have shown that reduplication is not the typical substitution pattern for children learning Mandarin in Singapore.

5.2 Tone analyses

5.2.1 Frequency of tone sequences

A heat map table of the tone sequences in the child forms in response to the nonword targets is presented in Table 5.3. This heat map will give us an idea of the proportion of occurrence of each tone sequence in response to the eight tone sequences of the nonwords. The main question is whether children's tone productions match those of the nonword targets, and if not, what kinds of substitute tone sequences the children make.

Table 5.3: Heat map table of frequency of occurrence of tone sequences in child forms, in response to nonwords

Legend: S1 = first syllable, S2 = second syllable, underlined text indicates highest proportion, darker shades indicate larger proportions

child forms of nonword targets		S2					S1 mean
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 0	
	Tone 1	0.03	0.16	0.11	0.15	0.09	0.11
	Tone 2	0.01	0.07	0.09	0.01	0.01	0.04
S1	Tone 3	0.10	0.02	0.00	0.01	0.00	0.03
	Tone 4	0.00	0.03	0.00	0.03	0.01	0.01
	S2 mean	0.03	0.07	0.07	0.05	0.03	

In general, T1 was the most prevalent in S1 of children's responses to the nonwords (as seen in the rightmost column 'S1 mean'), while T2 and T3 were the most common in S2 (as seen in the bottom row 'S2 mean). Specifically, children produced T1-2 (16%) most frequently, followed by T1-4 (15%), then T1-3 (10%). This is unsurprising as these three tone sequences were among the eight tone sequences (i.e., 12 of 32 nonword targets, 37.5%) in the nonword task. What is surprising is that of these three tone sequences, only T1-4 is a frequently occurring tone sequence by the CDI measure; T1-2 and T1-3 are infrequently occurring tone sequences.

5.2.2 Level of precision of tone sequences and patterns of substitutions

Next, the discrepancies between the frequency of occurrence of the target tone sequences in the child forms and the precision of these tone sequences are presented in the following Table 5.4.

Table 5.4: Frequency of occurrence and precision of child responses to nonword targets

Legend: Frequency of occurrence in child forms are taken from Table 5.3, table is ordered by the decreasing order of 'mean precision', grey cells indicate infrequent tone sequences based on CDI, underlined text indicates highest proportion in each category, **bold** text: referenced in text below.

nonword targets	child forms		
tone sequence	frequency of occurrence of tone sequence	mean precision relative to nonword targets	most common response/substitution
T3-1	0.10	0.72	T3-1
T2-3	0.09	0.60	T2-3
T1-4	0.15	0.57	T1-4
T1-3	0.11	0.51	T1-3
T1-2	0.16	0.49	T1-2
T2-2	0.07	0.41	T2-2
T4-2	0.03	0.17	**T1-2**
T4-4	0.03	0.16	**T1-4**

The eight tone sequences (✕ four nonword targets each) in the nonword task had been selected according to whether they were frequently or infrequently occurring tone sequences (four each) in the Singapore CDI.

In general, the most common response for the tone sequences matched, i.e., children were fairly accurate in producing the target tone sequences for all nonword targets, except for T4-2 and T4-4. In terms of frequency of nonword targets, three of the four frequent tone sequences also had the highest mean level of precision (T3-1, T2-3 and T1-4). The least frequent of the 'frequent tone sequences' (T4-4) was actually the least precise of the eight combinations: this was because T4-4 was often substituted with T1-4.

There is nonetheless a discrepancy between the frequency of occurrence in the child forms and precision. Children do not produce T3-1 most frequently but they are highly precise in producing T3-1 for the T3-1 nonword target. As mentioned previously, the three most common tone sequences in the child forms are T1-2, T1-4, then T1-3. The discrepancies between the frequencies of occurrence and precision for T1-2 and T1-4 is interesting: these sequences have the highest frequencies of occurrence but precision is only about 50%. T1-2 and T1-4 are highly frequent not only because both are among the target tone sequences in the nonwords, but they also substitute for the T4-2 and T4-4 nonwords, respectively (see **bold** text), which explains the fall in precision. T1-3 was also a frequent substitution tone sequence, which explains why these most common tone sequences had mid-range precision. These patterns of substitutions are discussed further at the end of this chapter.

5.3 Statistical analyses

To recapitulate, the research question addressed in the statistical analyses is to examine how, if at all, does the frequency of occurrence of segmental sequences and tone sequences influence the accuracy of child productions of newly encountered stimuli?

Singh et al. (2015) showed that children aged 2.5-3.5 years (close to the age of the children in this study) were more sensitive to tone mispronunciations, whereas children aged 4-5 years were more sensitive to consonant and vowel mispronunciations. Additionally, tone is a more salient feature than segments and has been more well-practised from a young age than segments. The results from the naturalistic observations also show that children have acquired most tone sequences and show high levels of precision for some of them. Thus, the frequent tone sequences are hypothesised to be more accurate than frequent segmental sequences. Nonwords with frequent segmental sequences and frequent tone sequences (FSFT; abbreviations will be revisited in the concluding chapter; as mentioned earlier, 'segmental sequences' are used here in place of 'consonant sequences' for better contrast of S for segments and T for tones) are hypothesised to be produced more accurately than those with infrequent segmental sequences and infrequent tone sequences (ISIT). For the remaining two categories, the following interaction was expected: stimuli with infrequent segmental sequences and frequent tone sequences (ISFT) were hypothesised to be targeted and produced more accurately than stimuli with frequent segmental sequences and infrequent tone sequences (FSIT), again because the frequent tone sequences of the former might facilitate production accuracy more than the frequent segmental sequences of the latter.

As the unit of analysis (segtone: segment accuracy vs. tone accuracy) is a binary variable, a series of binomial logistic regression analyses were performed (using the *glmer* function in the *lme4* package in R; Bates, Mächler, Bolker, & Walker, 2015; version 3.5.0.4, R Core Team, 2018) to evaluate the effect of different predictor variables on accuracy. Preliminary analyses of the raw scores for both segment accuracy and tone accuracy revealed higher scores in S2 than S1, which led to the observation that syllable position (pos: first vs. second) needed to be included as a factor in the analyses. The factors frequency of segmental sequences (segfreq: frequent vs. infrequent) and frequency of tone sequences (tonefreq: frequent vs. infrequent), as well as their interaction, were included as fixed effects. By-child and by-item variation were modelled with random intercepts. The initial model [accuracy ~ segtone * pos + (1 + segfreq * tonefreq | child) + (1 | item)] also contained random slopes for frequency of segmental sequences and tone sequences by child but it

failed to converge. Model reduction was performed by removing both slopes and convergence without singularity was obtained.

Following the advice of better statistical rigour of models using effect coding than treatment coding (see recent reviews by Brehm & Alday, 2022; Schad et al., 2020), effect coding was adopted for all (binary) fixed factors in the above models, where one level received a code of -1 and the other a code of 1 (see Table 5.5 below). This allowed effects to be interpreted with respect to the unweighted group mean rather than a particular base level. For all statistical analyses, the alpha level used for significance testing is $p < 0.05$. Likelihood ratio models have been built into logistic regression models (indicated by the p-values), hence no separate likelihood ratio models were run.

Table 5.5: Correspondence of codes with levels in effect coding for fixed effects models

Factor	Code: -1	Code: 1
unit of analysis 'segtone'	segment	tone
syllable position 'pos'	first	second
frequency of segmental sequences 'segfreq'	infrequent	frequent
frequency of tone sequences 'tonefreq'	infrequent	frequent

Violin plots were drawn to display the descriptive statistics of the variables.

The effects of the units of analysis and syllable position on total segment and tone accuracy are depicted in Figure 5.1 and Table 5.6.
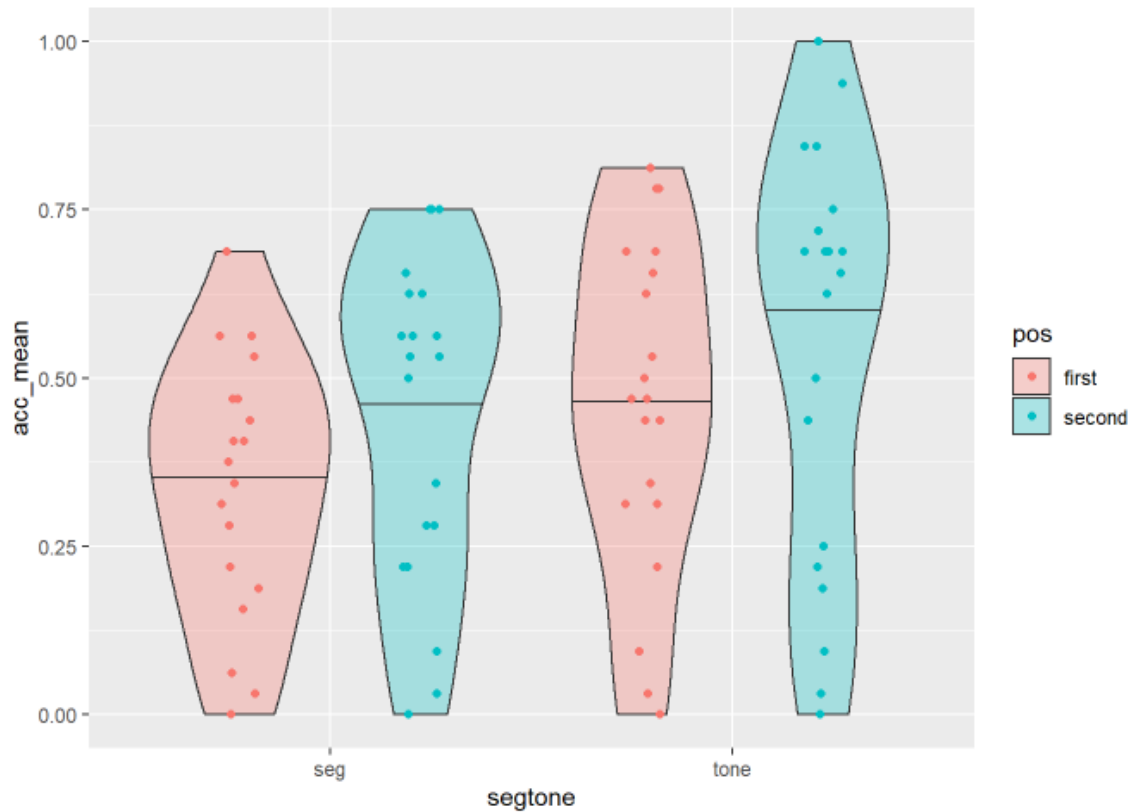
Figure 5.1: Mean accuracy of segment accuracy vs. tone accuracy by syllable position. Horizontal lines denote median.

Tone accuracy ($M = 0.50$, $SD = 0.50$) was significantly higher than segment accuracy ($M = 0.39$, $SD = 0.49$). Responses in S2 ($M = 0.49$, $SD = 0.50$) were significantly more accurate than those in S1 ($M = 0.40$, $SD = 0.49$).

Table 5.6: Logistic regression model for assessing the unit of analysis and syllable position on **total segment and tone accuracy**

Legend: segtone = unit of analysis (segment accuracy or tone accuracy), pos = syllable position (S1 or S2)

	Estimate	Std.Error	*z* value	Pr(>|z|)	
(Intercept)	-0.47	0.35	-1.37	0.17	
segtone	0.30	0.05	6.44	< .001	***
pos	0.26	0.05	5.55	< .001	***
segtone:pos	-0.03	0.05	-0.72	0.47	

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " "
glmer(accuracy ~ segtone * pos + (1 | child) + (1 | item), data=sample, family=binomial))

There was a significant main effect of 'segtone' (β = .30, z = 6.44, p < .001) and a significant main effect of syllable position (β = .26, z = 5.55, p < .001), reflecting the findings from the descriptive statistics: tone accuracy was significantly higher than segment accuracy and responses in S2 were significantly more accurate than those in S1. However, the interaction of these two factors was not significant.

The effects of frequency of segmental sequences, tone sequences and syllable position on **segment accuracy** are depicted in Figure 5.2 and Table 5.7.
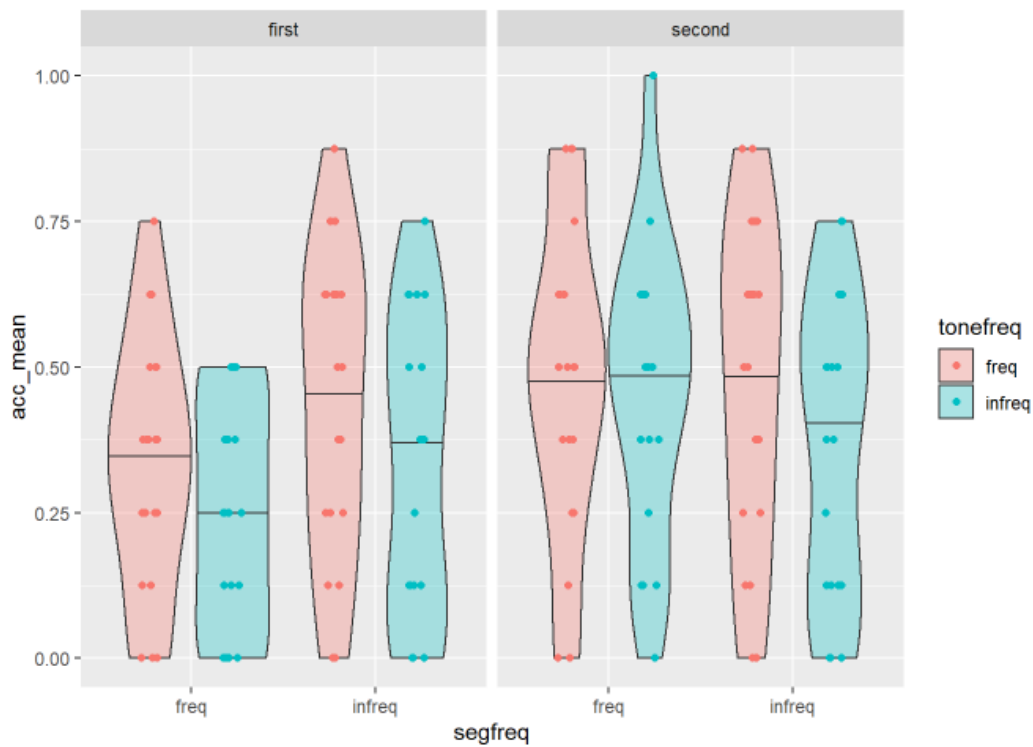

Figure 5.2: Mean accuracy of frequency of segmental sequences and frequency of tone sequences by syllable position on segment accuracy. Horizontal lines indicate median.

Table 5.7: Logistic regression model for assessing the frequency of segmental sequences, tone sequences and syllable position on **segment accuracy**

Legend: segfreq = frequency of segmental sequences, tonefreq = frequency of tone sequences, pos = syllable position (S1 or S2)

	Estimate	Std.Error	*z* value	Pr(>|z|)	
(Intercept)	-0.68	0.30	-2.25	0.02	*
segfreq	-0.06	0.11	-0.61	0.54	
tonefreq	0.20	0.11	1.88	0.06	.
pos	0.29	0.07	4.47	< 0.001	***
segfreq:tonefreq	-0.06	0.11	-0.61	0.54	
segfreq:pos	0.19	0.07	2.96	0.003	**
tonefreq:pos	-0.05	0.06	-0.80	0.42	
segfreq:tonefreq:pos	-0.05	0.06	-0.78	0.43	

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " "
glmer(accuracy ~ segfreq * tonefreq * pos + (1 | child) + (1 | item), data=sample_seg, family=binomial))

There was a significant main effect of syllable position (β = .29, z = 4.47, p < .001): S2 was significantly more accurately produced than S1. There was a significant interaction of frequency of segmental sequences and syllable position (β = .19, z = 2.96, p < .01) on segment accuracy: frequent segmental sequences were more accurately produced in S2 than in S1. None of the other factors or interactions made a significant contribution to segment accuracy. These findings are depicted in Figure 5.3 below.

Figure 5.3: Plot showing the effect of interaction between frequency of segmental sequences and syllable position on segment accuracy. Error bars indicate standard errors.

The effects of frequency of segmental sequences, tone sequences and syllable position on **tone accuracy** are depicted in Figure 5.4 and Table 5.8.
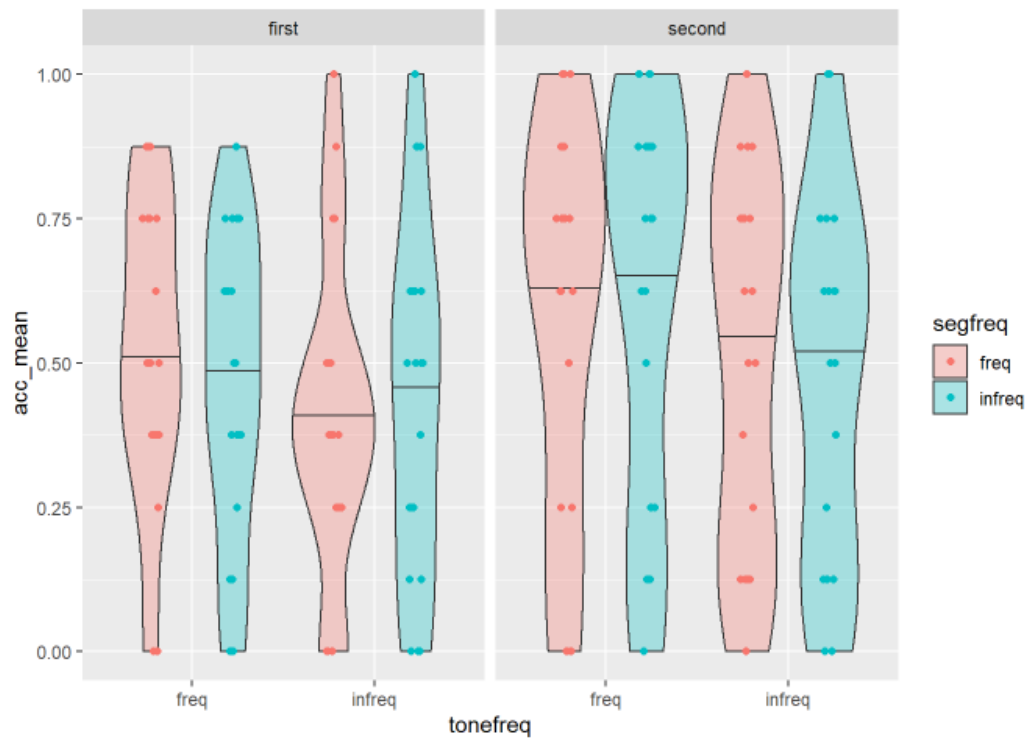
Figure 5.4: Mean accuracy of frequency of segmental sequences and frequency of tone sequences by syllable position on tone accuracy. Horizontal lines indicate median.

Table 5.8: Logistic regression model for assessing the frequency of segmental sequences, tone sequences and syllable position on **tone accuracy**

Legend: segfreq = frequency of segmental sequences, tonefreq = frequency of tone sequences, pos = syllable position (S1 or S2)

	Estimate	Std.Error	*z* value	Pr(>|z|)	
(Intercept)	-0.16	0.37	-0.43	0.67	
segfreq	0.04	0.11	0.32	0.75	
tonefreq	0.25	0.11	2.21	0.03	*
pos	0.24	0.07	3.58	< .001	***
segfreq:tonefreq	0.01	0.11	0.09	0.92	
segfreq:pos	-0.04	0.07	-0.53	0.60	
tonefreq:pos	0.06	0.07	0.84	0.40	
segfreq:tonefreq:pos	-0.10	0.07	-1.44	0.15	

Significance codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " "
glmer(accuracy ~ segfreq * tonefreq * pos + (1 | child) + (1 | item), data=sample_tone, family=binomial))

There was a significant main effect of tone frequency (β = .25, *z* = 2.21, *p* < .05) and a significant main effect of syllable position (β = .24, *z* = 3.58, *p* < .001). Frequent tone sequences were produced more accurately than infrequent tone sequences, and S2 was significantly more accurately produced than S1. None of the other factors or interactions made a significant contribution to tone accuracy. These findings are depicted in Figure 5.5.
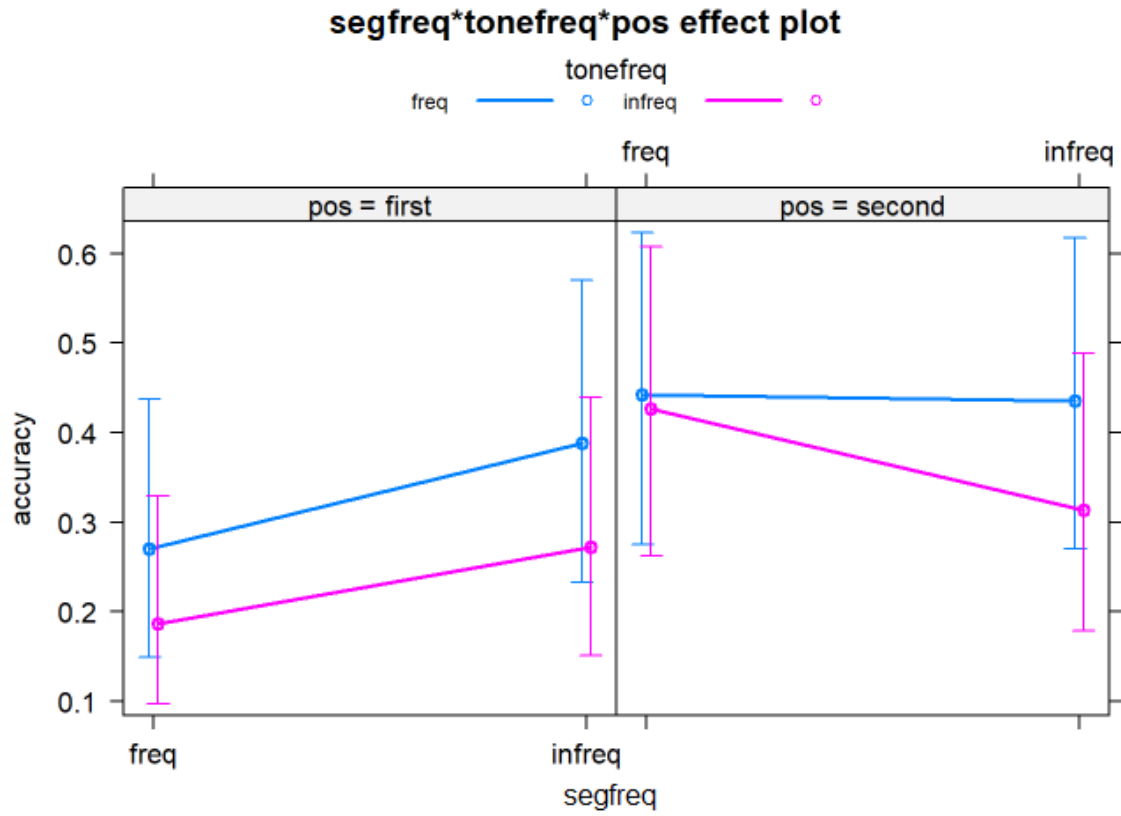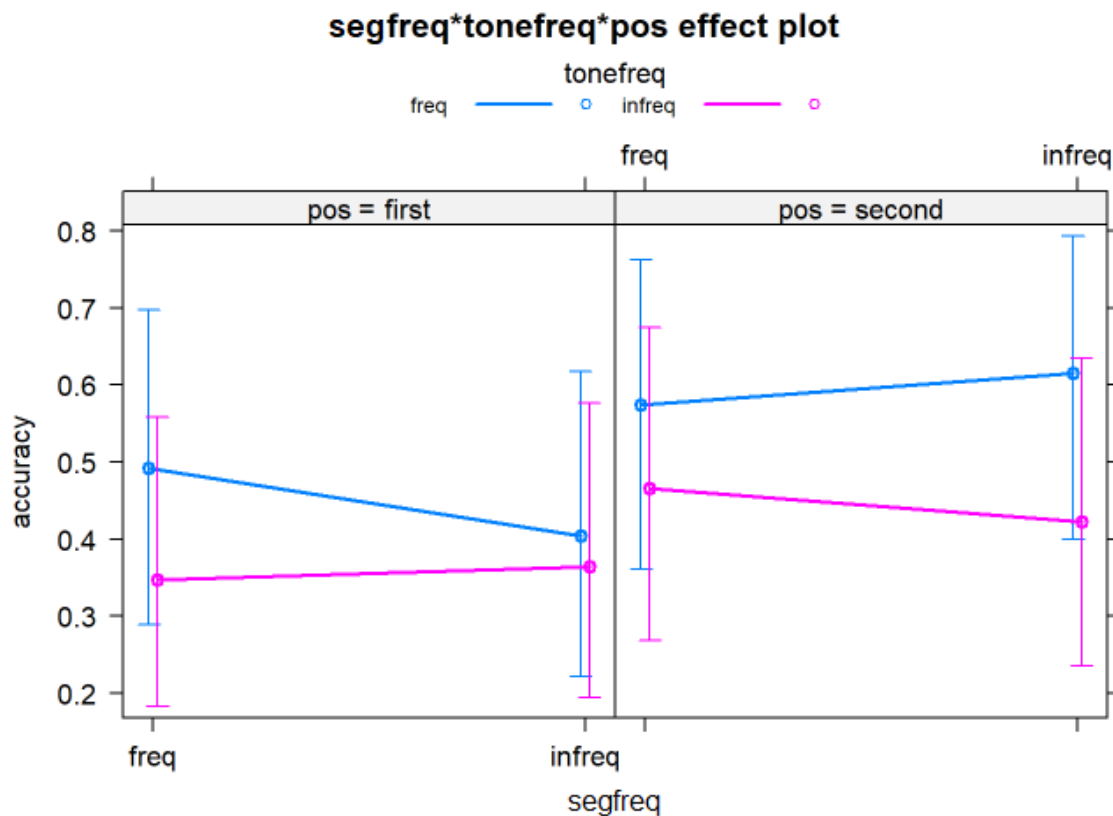
Figure 5.5: Plot showing the (lack of) interaction between frequency of segmental sequences and syllable position on tone accuracy. Error bars indicate standard errors.

The summary of findings from all the logistic regression analyses is presented in Table 5.9.

Table 5.9: Summary of findings from logistic regression analyses

analyses	total segment and tone accuracy	segment accuracy	tone accuracy
findings	- position: S2 > S1 - unit of analysis: tone > segment	- position: S2 > S1 - segfreq * pos interaction: freq + pos2	- position: S2 > S1 - tonefreq: freq > infreq

In sum, syllable position influenced all three analyses. In general, tone sequences were more accurately produced than segmental sequences. The frequent segmental sequences factor by itself did not influence segment accuracy but did when syllable position was included alongside it, whereas the frequent tone sequences factor by itself influenced tone accuracy.

5.4 Discussion

The research questions addressed in this chapter relate to the production and substitution patterns of the manner categories of word-initial consonants, word structures of the nonword responses and frequency and accuracy of tone sequences (descriptive analyses). Additionally, how the frequency of segmental sequences and tone sequences and interaction with syllable position would influence segment and tone accuracy in a nonword repetition task (statistical analyses) was also examined. Each of the analyses will be discussed in turn.

5.4.1 Moderate precision of segmental sequences and word structures

In terms of segmental sequences, the most common responses largely matched the manner categories of the targets, e.g., children were still producing FV-NV responses for an FV-NV target, they were not simply substituting infrequent segmental sequences with frequent ones. In terms of substitutions, fricatives were often substituted with plosives or omitted, and nasals and liquids were often omitted in word-initial onsets. This pattern of frequent omission for consonants in *nonwords* also follows the findings in children's productions of *words* (cf. Li & To, 2017). Words with nasal codas in S2 (half of the 32 nonword targets) were mostly produced as such; otherwise, they were omitted. The deletion of the final consonant has often been found in child productions in other languages (e.g., Macken, 1979/2013, on Spanish; Demuth & Johnson, 2003, on French).

In terms of word structures, the children were also fairly precise in producing responses that matched the word structures of the nonword targets. The nonwords were a 3:1 split between variegated and 'other' forms, which is similar to the distribution in frequency of variegated (61%) and 'other' (31%) words in the Singapore CDI; the remaining CDI distribution was 5% reduplicated and 3% consonant-repeated nonword targets. Variegated nonword targets were produced as variegated responses about 50% of the time, with 35% of the responses being 'other' (i.e., one consonant in the disyllabic nonword was omitted) and 6% of the responses being harmonised. 'Other' nonword targets were produced as 'other' about 70% of the time, and only about 10% each of the children's responses to these nonwords were harmonised or variegated.

As mentioned in the discussion of the patterns found in the naturalistic data in Chapter 4.1, in studies of child word production in other languages, harmony is usually the substitution of choice for variegated targets (e.g., Stoel-Gammon & Stemberger, 1994; Vihman, 1978), and reduplication and harmony are commonly used in response to challenging word structures like variegation (for example, as 'compensatory strategies': Menn, 1983; O'Seaghda et al., 2013). However, the Mandarin-learning children in the present experimental study produced variegated responses for variegated targets, even if the children substituted either or both consonants. This could be because of the proportion of words in the language that children hear: variegated words make up 61% of the words in the Singapore CDI and 57% of mothers' input (Table 4.5 in Chapter 4). Furthermore, the nonword repetition stimuli did not include consonant-repeated or reduplicated forms that are also present in the language, so children were exposed to a slightly higher percentage of variegation and 'other' words in the nonword repetition task than is typical. As a result, they may be stimulated to maintain variegation in their child forms when repeating the nonwords, within the limits of the nonword repetition task.

5.4.2 Systematic patterns in tone substitutions

The earlier sections revealed the prominence of the T1-*x* tone sequences in response to a nonword repetition task, not least because the task included many nonwords with T1-*x* tone sequences. The most frequent patterns of tone sequence substitutions for the nonword repetition task are examined further (Table 5.10).

Table 5.10: Observable patterns in children's most frequent patterns of tone sequence substitutions

Legend: grey cells indicate infrequent tone sequences based on CDI

nonword targets	most frequent patterns in child substitutions	
tone sequence	**T1-*x* (excluding T1-0)**	**T1-0**
T1-2	T1-3	T1-0
T1-3	T1-2	
T1-4	T1-3	T1-0
T2-2	T1-2	T1-0
T2-3	use of T1-3 = T1-2	
T3-1	T1-1	
T4-2	T1-2	T1-0
T4-4	T1-4	T1-0

In analysing these substitutions, there are two findings of note. First, all the resultant tone sequences have T1 in S1. Given that three out of the eight tone sequences (12 out of 32 nonword targets, 37.5%) in the nonword task have T1 in S1, it seems that this within-task frequency and likely the acoustic salience of T1 itself led the resultant child substitutions to have T1 overwhelmingly produced in S1. Additionally, the prevalence of T1-*x* tone sequences is also recorded in everyday speech and has been a well-practiced tone sequence since the babbling period (cf. Lou et al., 2018; see Chapter 4.2).

Next, most of these patterns occur infrequently in the language (see grey cells, calculated based on the CDI): only T1-1 and T1-4 are frequent in the language, the remaining tone sequences are infrequent. Thus, it is not that children are changing an infrequent tone sequence to a frequent tone sequence – in fact, children are switching some of the frequent tone sequences in the task to *infrequent* ones to have T1 in S1. This could be explained as a whole-word effect of the salience of T1: the acoustic salience of T1 takes precedence over the frequency of occurrence of the tone sequence in the language (see Chapter 4.2 and final Chapter 6).

Within the T1-*x* substitutions in the nonword repetition task, two patterns can be observed: 1) T1 is the substituted tone of choice in S1; and 2) T1-0 is a general substitution tone sequence: similar to its usage as has been discussed in Chapter 4.2, T1-0 is an adapted 'tone template' (cf. Mok & Lee, 2018).

*5.4.2.1 T1-*x* pattern*

From Table 5.10, the most frequently occurring tone sequences children produced were fairly accurate and largely matched the target tone sequence (e.g., children produced T1-3 responses most frequently in response to the T1-3 nonwords), except the T4-2 and T4-4 targets (**bold** text in Table 5.4). For these, T4 in S1 was substituted with T1. This finding of T4 being substituted with T1 corresponds to the finding in the naturalistic data (Chapter 4) where T4 in S1 was especially likely to be produced as T1. For the nonwords, it seems that this T4 → T1 substitution in S1 carried over to the nonword targets with T2 and T3 in S1 as well: the next most frequent responses for the T2-2, T2-3 and T3-1 targets were T1-2, T1-3 and T1-1 respectively. T2-3 targets were additionally substituted with T1-2 as the next most frequent response.

Even the T1-*x* nonword targets themselves were substituted with another T1-*x* tone sequence, specifically, either T1-2 or T1-3. As mentioned, T1-*x* tone sequences were overly represented in the nonword task, including T1-2 and T1-3, thus the presence of these tone sequences may have been relatively salient to the children for the duration of the nonword task.

5.4.2.2 T1-0 tone template

It is interesting that T1-0 is also a frequent substitution response to the tone sequences; it seems to be a general substitute tone sequence, as there is no specific pattern that can be discerned among the tone sequences that result in the T1-0 substitution. Furthermore, it is an infrequently occurring tone sequence in the language: T1-0 makes up only 4% of the words in the Singapore CDI; T1-0 in mothers' input (from Chapter 4) also made up only 4%. From the results of the naturalistic observations in Chapter 4, T1-0 was the *most frequently* produced tone sequence for words (17%), but it was also the one *least precisely* produced. This is because it was the *de facto* substitution tone sequence for other words, as with the nonwords in this chapter.

For the nonword repetition task, the proportion of occurrence of T1-0 in child forms was 9% (less than half as much as the most frequently occurring T1-2, 16%), with precision being 0 since there were no T1-0 targets. It seems likely that the arguments laid out in Chapter 3 for the prevalence of T1-0 applies here too: T1-0 is an adapted tone template that children most frequently produce, even in response to unfamiliar stimuli. This is because the tone sequence is a falling tone contour that is fairly effortless to produce, the T1-0 tone sequence might reflect cross-linguistic transfer from English (cf. Mok & Lee, 2018) and it is also a well-practiced tone sequence (cf. Lou et al., 2018).

5.4.3 Independence of segment and tone; tone accuracy higher than segment accuracy

The most striking finding from the statistical analyses was that segment and tone accuracy were independent of one another. The frequency of segmental sequences had no effect on tone accuracy; nor did the frequency of tone sequences have an effect on segment accuracy. Surprisingly, syllable position played an important role in all three analyses: whichever the analysis, there was a main effect of syllable position, such that responses in S2 were more accurately produced than those in S1. This is probably a recency effect (cf. Jones et al., 2004), in that the segment in the second position is the most recent element.

The frequency of segmental sequences did not by itself have an effect on segment accuracy: children performed equivalently on the nonwords regardless of the frequency of segments. This is interesting, as the effect of frequency is pervasive in language acquisition generally (see Ambridge et al., 2015). In this study, frequency of segmental sequences was calculated differently from what has previously been done, which could perhaps explain the unexpected finding. Previously, controls for frequency in nonword repetition tasks have either used a subjective judgement measure of wordlikeness (e.g., Gathercole et al., 1991) or a more objective method of calculating positional segment frequencies and biphone frequencies to determine the phonotactic probability of a word (e.g., Vitevitch & Luce, 2005). In this study, the segmental sequences were not contiguous, they were the syllable-initial segments of a disyllabic word. Furthermore, these previous counts have also relied on adult judgement or adult word databases (cf. Sosa, 2008; Sosa & Stoel-Gammon, 2012). The control for the manner categories of disyllabic words in the frequent and infrequent segmental sequences using a child corpus (Singapore CDI) has not been done before, so this may be a reason for this difference. Future studies should explore this frequency effect further.

The descriptive analyses of the effect of frequency of segmental sequences on segment accuracy support this lack of effect on segment accuracy from the statistical analyses: there did not seem to be any difference for children in producing frequent or infrequent segmental sequences. Children did not find the frequent segmental sequence any easier to produce than the infrequent segmental sequence; there is no clear-cut finding of infrequent sequences being substituted with frequent sequences (cf. Table 5.1). Individual consonants were produced, substituted or omitted – plosives were produced as plosives, fricatives were produced as fricatives or substituted with plosives, and nasals and liquids may be omitted – regardless of the frequency of segmental sequences.

However, there was an effect of frequency of segmental sequences on segment accuracy when syllable position was also a factor, namely, frequent segmental sequences were produced more accurately in S2 than S1. Here, the recency effect influenced children's production of frequent sequences in S2.

Comparing the findings on segment accuracy in the present study with the only previous study of nonword repetition in Mandarin, recall that Wang and Huang (2016) commented that the children suspected of DLD and age-matched controls scored better for high-wordlike nonwords (cf. Gathercole, 1991) than for low-wordlike nonwords. As mentioned in section 3.2.1, some specific nonword syllables used in the nonword task in the present study were

said to be less wordlike than others; however, there was no difference in performance between the more and less wordlike nonword syllables. Results from the wordlikeness judgement task with adults also confirmed that all the nonwords were comparable in terms of (non)wordlikeness. The frequent and infrequent segmental sequences of nonword syllables in the present study may also be equated with high and low wordlikeness, respectively; however, as revealed by the earlier findings, there was also no difference in performance between frequent and infrequent segmental sequences.

While there was no frequency effect of segmental sequences on segment accuracy, there was a frequency effect of *tone* sequences on *tone* accuracy: frequent tone sequences were produced more accurately than infrequent tone sequences. These results from the statistical analyses parallel the findings from the earlier descriptive analyses (recall Table 5.4): the top three most precise tone sequences, T3-1, T2-3 and T1-4, were considered frequent in the Singapore CDI. However, children produced T1-*x* tone sequences most frequently, many of which were considered *infrequent* in the Singapore CDI. There is thus a three-way discrepancy: what is considered frequent in the language (as measured by the Singapore CDI) may not be the same as what is frequently produced by the child (cf. Menn, 2013, Ota, 2013; see Tables 5.3 for the most frequently produced tone sequences), which may furthermore not correspond to their precision level.

Comparing the findings of tone accuracy in the present study with Wang and Huang (2016), the five disyllabic nonwords in the latter had all been produced in the T1-1 tone sequence. Thus, tone was not a factor for their study; the influence of tone sequence on nonword repetition tasks has not been studied till now. In this study, frequency of tone sequences was calculated from the frequency count in the language, by means of the words in the Singapore CDI. As mentioned earlier, frequent tone sequences were produced more accurately than infrequent tone sequences.

Nonetheless, neither frequency of segmental sequences nor any of the two-way and three-way interactions contributed to tone accuracy. This concurs with research that shows that tone may be processed separately from segments (cf. toneme in revised TRACE model, Ye & Connine, 1999 or the different neural processing patterns in the brains of people learning tone and non-tone languages, Gandour et al., 2000; Klein et al., 2001; Wong et al., 2004). Chen (J.-Y., 1999, 2000) has also suggested that Mandarin lexical tone production is analogous to English stress production, where research on speech production (Levelt, 1993) has shown that English stress is planned earlier than segmental information: the stress

pattern provides the phonological frame for the word and the segmental information is then inserted into the frame. Similarly, tone provides the 'envelope' for which segments are filled in. Thus, neither segment nor tone influences the other.

Comparing the two elements of accuracy, tone accuracy was significantly higher than segment accuracy. This supports previous findings that tone errors tended to occur less frequently than segment errors in adult speech production (Chen, J.-Y., 1999). Fernald (1989) explained that tone has an advantage over segments because it was part of vocal prosody that children have been attentive to from young and could help the children learn words early on. Furthermore, researchers who examined the segmental and tone error patterns in *single word* production have proposed that there is a dissociation in terms of errors depending on the location of the error (Chen, J.-Y., 1999, 2000; Wan & Jaeger, 1999). In their error analyses on adult speech production, they claim that if the segmental error occurs within a word, the tone may retain its accuracy, but if the segments across the whole word are erroneously produced, the tone usually changes along with the segments. The findings from the nonword repetition task of *disyllabic words* and thus *tone sequences* produced by the children in this study do not support this claim – many children still produced tone sequences accurately even if they made errors in the segments of the disyllabic nonword target; some children were also accurate in their segmental sequences while they made tone errors, although this latter pattern happened to a lesser extent (given the fact that segment errors still occurred more often than tone errors). Perhaps there is a difference between child and adult performance: for children, tone errors in tone sequences are independent of segmental errors, no matter where the segmental error is.

In the next and final chapter, the findings from both naturalistic and experimental studies are integrated; the evidence (or lack thereof) of templates in children's production of Mandarin words and nonwords and how the whole-word approach may be applied to Mandarin are discussed.

Chapter 6 – Phonological development in Mandarin

What do the findings regarding children's segment and tone productions and patterns tell us about phonological development in Mandarin? The thesis combined observational and experimental approaches to examine children's phonological development in Mandarin. In the naturalistic observational study, children's speech was observed in a play setting monthly over six months and the words they targeted and produced were analysed alongside their mothers' input. In the experimental nonword repetition task, nonwords were designed to elicit children's imitation responses to wholly unfamiliar word forms. Across both paradigms, production patterns and error patterns of segmental sequences and tone sequences were noted and any templates identified.

This chapter summarises the major findings of the study in relation to the research questions outlined in section 1.4. An overview of the research questions is provided, followed by integration of the findings from the naturalistic and experimental tasks. The whole-word approach is then discussed with respect to how it may apply to Mandarin, in terms of identifying and tracking the rise and fall in templatic usage.

6.1 Overview of research questions and integrating naturalistic and experimental findings

Three major research questions have been addressed in this thesis, as delineated in Chapter 1 (section 1.4):

1) Is there overlap among the consonant sequences and tone sequences children hear, target and produce?
2) Is there evidence of templatic usage in the children's segment and tone productions?
3) How, if at all, does the frequency of occurrence of segmental sequences and tone sequences influence the accuracy of child productions of newly encountered stimuli?

The design of the study combined naturalistic and experimental paradigms. Four children aged 20 months were observed over a period of 6 months and their segmental and tone patterns in their cumulative speech were analysed. Twenty children aged 24 months were then tested on an experimental task of nonword repetition, designed to test the effects of the frequency of occurrence of both segmental sequences and tone sequences. Naturalistic observations provide insight into the children's spontaneous production patterns in everyday life. The experimental task then complements those observations by informing us as to how

children may apply the same patterns observed in their everyday interactions in repeating novel words, that is, words that they have never encountered before.

The findings were that the patterns of segments and tones children produced in the naturalistic observations were also reflected in the experimental task, that some of these patterns were different from what is found in the input, and that segment and tone were independently represented. In general, children were more accurate in tone sequences than segmental sequences, regardless of the lexical nature of the stimuli and the frequencies of occurrence of the segmental sequences and tone sequences in the language. A more detailed discussion of the segment and tone findings across the two types of production paradigms (naturalistic speech and nonword repetition responses) follows.

6.1.1 Segmental analyses

The productions and substitution patterns of consonant sequences and word structures in the words of children learning Mandarin were examined first. The children produced /t/-/t/ the most and each had their own other favoured consonant sequences. Their use of /t/ was also often substituted for the more challenging fricative and affricate sounds that they had not yet mastered (cf. Ma et al., 2022), thus even though they were frequently used, they were not considered templatic. When comparing across languages, while there was a higher incidence of fricative-initial words in Mandarin relative to other languages and the proportion of fricative-initial words mothers produced also matched the higher proportion in the language, children still produced plosive-initial words the most often (both accurately and as substitutes for fricatives), corresponding to findings from cross-linguistic studies (cf. Vihman et al., 1994). This likely reflects the articulatory ease of plosives relative to fricatives (cf. MacNeilage & Davis, 1990). A similar pattern of production of plosives and fricatives was also found in the experimental nonword repetition task.

For word structures, the corresponding high proportion of variegation and negligible proportion of consonant-repetition/harmony across both production tasks (naturalistic speech and experimental nonword repetition) mirrored their respective proportions in the language, parental input and within-target frequencies. There were some reduplications in the naturalistic data but none in the nonword repetition task, no doubt as a result of task demands: reduplicated words are present in the language and in the everyday speech of mothers and children but none were included among the targets in the nonword repetition task. In the Singapore CDI 18 of 372 disyllabic words (5%) are reduplicated, with a prominent category being kinship terms; these are also a source of several tone templates.

In the nonword repetition task the ratio of variegated to 'other' nonword targets was 3:1, which is similar to their proportions of occurrence in the language. However, the task excluded the reduplicated and consonant-repeated word structures also found in the language. This means that variegated and 'other' words are slightly over-represented in the task relative to the language, which may have primed children to produce more variegated and 'other' responses than is typical within the limits of the nonword repetition task. However, even in the naturalistic data the children rarely adopted reduplication and harmony as substitutions, so it is understandable that children in the nonword repetition study failed to produce reduplicated forms in response to the variegated and 'other' nonwords.

Harmony is a mechanism that children often employ to tackle variegated words. These are said to be the most challenging type of word for a child to produce as they involve the production of two supraglottal consonants. Yet the children in this study did not follow this pattern. Vihman et al. (2022) propose that the overall simple syllable structure of East Asian languages (specifically, Japanese and Mandarin) enable children to maintain variegation in their productions at a higher rate than the children learning the other languages. Japanese and Mandarin need only a 1:1 syllable-to-word ratio to produce disyllabic words, whereas languages like English, Finnish and French need between a 1.2 to 1.5:1 ratio. That is, for every 10 disyllabic words in Japanese or Mandarin these would contain on average 10 different syllables, whereas in the other languages the same number of words would, on average, contain more (between 12 and 15) different syllables. Vihman et al. conclude that the smaller inventory of syllables needed to learn and produce word forms may make it easier for children learning Japanese and Mandarin to remember word patterns. leading children to attempt more challenging variegated word forms. This could also explain why frequent and infrequent segmental sequences in the nonword repetition task were produced as such; practice with the simple syllable structure may have allowed children to produce even the infrequent segmental sequences rather than needing to substitute more frequent segmental sequences.

Nonetheless, there were differences between the findings of a recent systematic study of Mandarin word production (Lou, 2020) and the present study. Lou found more reduplication and less variegation than was found in the present study and a wider range of types of reduplication (as previously presented in Table 4.9). As discussed in Chapter 4.1, the difference is likely due to the higher lexical level of the children in the present study and their relatively greater experience of word shapes and greater practice with production than with Lou's children, This may also relate to the more advanced word types mothers use with their children. This likely resulted in the higher proportion of variegation and lower proportion of

reduplication found in both mothers' and children's productions in the present study relative to Lou (2020). This may also explain the higher proportion of 'other' than of reduplicated words produced by the children in response to variegated targets here. Finally, the language variety and cultural differences in the reduplicated input children receive from their mothers likely also partially account for the differences between the two datasets, as was discussed.

6.1.2 Tone analyses

Next, tone was examined. The present study is the first to directly analyse the productions and substitutions of *tone sequences* in the words of children learning Mandarin. Use of systematic patterns or tone templates amongst the words was also identified. There have been mixed findings about the mastery of *single tone* acquisition by age 2, mostly as a result of inconsistent definitions of mastery/acquisition as well as the use of different experimental paradigms. In this study, adapting from Mok et al.'s (2020) proposed definition of 'above-chance level' of accuracy in monosyllabic Cantonese stimuli to suit disyllabic Mandarin tone sequences, children may have acquired most of the tone sequences by age 2. Nevertheless, there was a protracted period of precision for *tone sequences.*

There was a mismatch in terms of the tone productions children heard, targeted and produced: the three most frequently produced tone sequences across all data sources are summarised in Table 6.1.

Table 6.1: Top three most frequently produced tone sequences by data source (all sources are from the naturalistic data except 'child nonword forms')

Legend: grey cells indicate infrequent tone sequences based on Singapore CDI

source	most frequent	2nd most frequent	3rd most frequent
mothers' input	T2-3	T3-4	T4-4
child targets	T2-3	T3-1	T1-1/T4-4
child word forms	T1-0	T1-1	T1-4
child kinship term forms	T1-0	T3-2	T1-1
child nonword forms	T1-2	T1-4	T1-3

There are two patterns of note in Table 6.1. First, for the most part, children are targeting tone sequences similar to those that their mothers are producing. T2-3 is a common tone sequence across all of them, and so are some tone sequences with T3 in S1 and T1 and T4 in S2. All of these tone sequences also occur frequently in the language (based on the Singapore CDI). However, there was no complete overlap between mothers' input, child

targets and child forms. Second, there is also a commonality across the most frequently produced child forms of words and nonwords: with the exception of T3-2 in kinship terms, all are T1-*x* sequences, indicating some templatic usage of T1-*x* tone sequences.

Of the seven tone sequences observed among the child forms, about half were frequent and half infrequent in the language (based on the CDI), regardless of lexicality. Additionally, half the overwhelmingly T1-*x* sequences were considered infrequent in the language: T1-2, T1-3, and T1-0. However, as mentioned, *frequency* of use in the child forms proved to be independent of precision in production. A further caveat needs to be added: only eight tone sequences were included in the nonword repetition task, a restricted range of potential tone sequences in targets compared to the range possible. Additionally, tone sequences with T0 in S2 had been intentionally excluded from the initial stimuli creation phase because prior research had indicated a lack of use of T0 in S2 in Singapore (Shang & Zhao, 2013, 2017). Words in the CDI also show that T0 in S2 did not occur frequently. In the CDI, three types of words have T0 in S2: 1) reduplicated words (mostly kinship terms); 2) nouns that end with the affix *tsɹ0* (which occurred frequently among Fu Zhen's words), some of which can be omitted (e.g., *pei1tsɹ0* 'cup' may simply be *pei1*; even though disyllabic words make up the majority of Mandarin words, many words and concepts can be expressed using just one syllable), and 3) grammatical words (e.g., *li0* often used with 'here' or 'there', *tʂə0* and *lə0* indicate participle, *tə0* indicates possessive, *kə0* and *mən0* indicate a count of items and people respectively, *mə0* is often used with 'how' or 'so', among others). Thus, most of the words with T0 in S2 are kinship terms and grammatical words, which occur less often than other lexical content words. In examining the production data in this study, there was some evidence of use of T0 in S2 in Singapore. Mothers produced kinship terms in various tone sequences, with some producing T0 in S2 while others produced a full tone. These two categories – nouns that end with affix *tsɹ0* and grammatical words – are said to be more often produced with a full tone in Singapore Mandarin (Shang & Zhao, 2013, 2017). An examination of how mothers produced these words in this study revealed a similar pattern of variation in the tone sequences as in kinship terms: sometimes mothers produced these grammatical words as T0, sometimes as full tones. Thus, T0 in S2 might be present in Singapore Mandarin, but more inconsistently, and with more variability, than is evident from the CDI or than previously reported by Shang and Zhao (2013, 2017).

For the nonword repetition task, the frequent production of the infrequently occurring T1-2 and T1-3 sequences may have been influenced by the fact that they were among the tone sequence targets in the nonword repetition task. Additionally, both sequences also substituted for others, especially for the T4-2 and T4-3 nonword targets. Thus, while the

T1-2 and T1-3 tone sequences may be infrequent simply because there are fewer words with these tones in the CDI, children gain a lot of practice in producing T1 in S1 due to its ease of production and occurrence in babble (cf. Lou et al., 2018; see further discussion in section 6.2.1 below). This also explains why T1-1 and T1-0 could be identified as 'tone templates' in the children's speech across both production paradigms, an extension of the original concept of the 'phonological template' (Vihman & Croft, 2007; Vihman, 2019).

On the other hand, children may have had less practice in producing the sequences taken here to be frequent, based on the CDI). There is a potential tension between what the children *hear* most (in the input or as estimated by the CDI) and what they *produce* most (in the child forms). These two aspects end up exerting opposing influences on accuracy of tone production: the frequently heard tone sequences are not well-practised, while the practised sequences are infrequently heard in the input, though they may be heard in their own production. As a result, no effect of frequency as measured in this study could be detected.

Recall that the most precisely produced tone sequences in the nonword repetition task (from Table 5.4) are T3-1, T2-3 and T1-4. From Table 6.1 above, T3-1 and T2-3 are among most *targeted* tone sequences in the naturalistic study. However, while the children targeted these tone sequences the most frequently, they did not produce these tone sequences as such. The forms produced were substituted with other tone sequences. It seems that in naturalistic productions, children did not spontaneously produce these tone sequences, but when encountering unfamiliar stimuli with the same tone sequences, they were able to imitate them. There may thus be a difference between long-term retrieval and short-term recall for certain tone sequences: in long-term retrieval, when children are unable to accurately recall the tone sequence, they produce their default T1-0 or other T1-*x* tone sequences instead.

Specifically, T1-1 appears to be a tone template that children select and adapt to in general, but especially for kinship terms. Kinship terms are highly relevant in a child's life and many of them have the sequence T1-1; kinship terms that are not T1-1 are adapted to the pattern. Furthermore, T1-1 is a tone sequence that involves repetition of the same acoustically salient tone, which makes it easy to produce. Meanwhile, T1-0 may be an adapted tone template in general for both words and nonwords, and a tone template children select and adapt in producing the kinship terms. T1-0 resembles the trochaic stress in English, which is widely spoken in the children's surroundings; it may reflect cross-linguistic transfer of English to Mandarin (cf. Mok & Lee, 2018). Articulatorily, it also constitutes a high contour that falls. In general, T4 occurs less often in S1. It is more common in S2, as the tone has a falling contour and is more effortlessly produced in S2. Thus, the T4-*x* tone sequences throughout

the present study (both naturalistic observations and experimental nonword repetition task) were overwhelmingly produced as T1-*x* tone sequences.

Finally, the category of kinship terms also included a third pattern, T3-2. In Lou (2020), parents and children did not produce T3 or T3-2 frequently at either the 4wp or the 25wp, likely due to T3's more challenging acoustic contour. Nonetheless, as mentioned in Chapter 2, the final rise of T3 (the **4** in the 214 tone contour, Chao, 1930 notation) may be realised only in isolation or at the end of the sentence (Chen, C. Y., 1984; Singh & Fu, 2016). The beginning fall portion of T3 (21) appears elsewhere (Chen, C. Y., 1984; Wang & Li, 1967). This makes T3-2 a relatively effortless tone contour to produce: the slight 2-1 fall of T3 leads to the 2-5 rise of T2. That is, the low F0 offset in T3 transits seamlessly to the low onset in T2 (cf. Wong, 2013). The prevalent usage of T3-2 may potentially be unique to Singapore as a result of reciprocal cross-linguistic transfer and language mixing between Cantonese and Mandarin as used in Singapore (unique because the resultant tone combination of T3-2 is different from the proposed origin in Cantonese, which is more similar to the Mandarin T3-**1**, as described in the Cantonese/Mandarin discussion in section 4.2.3.2.3). As a result, T3 is a relatively more frequently occurring tone for the children in Singapore than for the children in Lou's study. A multilingual tone environment brings about tone diversity and tone mixing. Perhaps a comparable set of analyses of language mixing in Hong Kong or Malaysia (where there is also the use in parallel of Cantonese and Mandarin, though Mandarin may be used to a lesser extent; additionally, the official language of Malaysia is Malay, adding yet another language to the mix) would tell us more about whether this pattern is indeed unique to Singapore.

6.1.3 Frequency effects on segment accuracy and tone accuracy

The influence of frequency on productions of novel stimuli was explored further with respect to how frequencies of segmental sequences and tone sequences might have influenced accuracy in children's productions of newly encountered words. Segment and tone proved to be independent: the frequency of segmental sequences did not influence tone accuracy, nor did the frequency of tone sequences influence segment accuracy. Frequency of segmental sequences was not significant until syllable position was included in the analysis of segment accuracy (frequent segmental sequences were more accurately produced in S2 than in S1); however, there was a significant effect of tone frequency on tone accuracy.

The analysis of frequency effects of the segmental sequences and tone sequences within the words produced across *both* production paradigms are discussed next.

In this section, the following abbreviations of the four frequency categories are used (first referenced in section 5.3):

- FSFT: stimuli with frequent segmental sequences and frequent tone sequences
- FSIT: stimuli with frequent segmental sequences and infrequent tone sequences
- ISFT: stimuli with infrequent segmental sequences and frequent tone sequences
- ISIT: stimuli with infrequent segmental sequences and infrequent tone sequences

Across all types of stimuli regardless of lexicality, a frequency effect was expected, i.e., stimuli with both frequent segmental sequences and frequent tone sequences (FSFT) were expected to be targeted and produced the most often and words with both infrequent segmental sequences and infrequent tone sequences (ISIT) were expected to be targeted and produced the least. For the remaining two categories, the following interaction was expected: stimuli with infrequent segmental sequences and frequent tone sequences (ISFT) were expected to be targeted and produced more than stimuli with frequent segmental sequences and infrequent tone sequences (FSIT). This is because Singh et al. (2015) showed that children aged 2.5-3.5 years (which is close to the age of the children in this study) were more sensitive to tones, whereas children aged 4 -5 years were more sensitive to consonants and vowels. Thus, children should have produced the *frequently* occurring tone sequences (as measured by the CDI) more frequently. However, the opposite pattern was found: children produced the *infrequently* occurring tone sequences more frequently.

Table 6.2 displays the distribution of the four frequency categories across the various modes of vocal production elicited in this study, namely, the combined spontaneous and imitated words in the naturalistic observations (Chapters 4) and children's responses to the experimental nonword repetition task (Chapter 5).

Table 6.2: Distribution of four frequency categories across data sources

Legend: Grey cells indicate intentional design of even distribution of target nonwords by frequency category

	mothers	naturalistic		experimental nonwords	
	input	target	child form	target	child form*
FSFT	0.49	0.52	0.42	0.25	0.24
FSIT	0.17	0.14	0.27	0.25	0.27
ISFT	0.23	0.24	0.17	0.25	0.20
ISIT	0.12	0.10	0.13	0.25	0.20

* Note: The proportions of child forms for the nonwords do not add up to 1 because miscellaneous or one-syllable responses (<10%) are not included in this table.

The proportions of the frequency categories across mothers' input and targets are the same, and in the expected pattern: FSFT stimuli had the largest proportion (about half) and ISIT stimuli had the smallest (about 10%). Stimuli with frequent tone combinations (FSFT and ISFT) formed the majority across these three data sources. This count corresponds with the tone combinations of mothers' input and child targets as discussed in Chapter 4 and in Table 6.1: all the tones that mothers produced and children targeted were frequently occurring tone sequences.

However, the pattern for *child forms* differed from mothers' input and the targets: the proportions of the FSFT and ISIT were the same as mothers' input and the targets, but FSIT stimuli were produced more often than ISFT, the reverse pattern of mothers' input and the (naturalistic) targets. Comparing between the targets and child forms, children were making substitutions to the words they hear: they **targeted** the words with frequent segmental sequence (and infrequent tone sequence, FSIT) less often but ***produced*** them twice as often. This count also corresponds with the tone sequence of child forms discussed in Chapter 4.2 and in Table 6.1: T1-0 is an infrequently occurring tone sequence in the language but children produced it the most, followed by T1-1 and T1-4 (which are frequently occurring in the language).

The children's responses to the experimentally manipulated task of nonword repetition produced yet another pattern of results. For this task, the proportions of the targets were artificially manipulated to be equally prevalent in each of the frequency categories. Children produced frequent segmental sequences (FSFT and FSIT) more than those with infrequent segmental sequences (ISFT and ISIT), but contrary to expectations again, the proportion of FSIT responses was greater than that of ISFT responses.

Throughout both naturalistic and experimental tasks, frequency of occurrence in the language has been shown not to equate directly with the frequency within the child's repertoire of words (cf. Menn, 2013; Ota, 2013); similarly, within-task frequencies may also influence children's productions for the duration of the task. As seen from the naturalistic (Chapter 4) and experimental data (Chapter 5), some of the most frequently produced segment and tone sequences were idiosyncratic child patterns and templates and not all were among the most frequently occurring in Mandarin. This is crossed with yet another factor, that of *precision* (i.e., the degree of match of child form to target): some of the frequently produced segment and tone sequences were not precise at all and were used mostly as substitutions, for example, some of the T1-*x* tone sequences. As mentioned in Chapter 4.2, T1-*x* tone sequences have been well-practiced since the babbling stage (cf. Lou et al., 2018). The patterns of use and prevalence of the T1-*x* tone sequences in the present study suggest templatic usage; these T1-*x* tone sequences are further discussed in section 6.2.1.

6.2 The whole-word approach to Mandarin

The goal of the thesis was to examine whether the whole-word approach of studying children's early speech forms may be applied to Mandarin and what form it might take. The basis of the whole-word approach is that children learn word forms; knowledge of sound categories emerges from knowledge of those forms (Vihman, 2017). Evidence of this comes from the observation that children's early word forms are 'holistic', quite accurate, and closely similar to salient aspects of the child's own phonological repertoire (Vihman, 2017). The salient aspects may affect memory for non-salient aspects. As children produce more and more words, routines and patterns may begin to emerge. The emergence of these patterns, termed 'phonological templates' (Vihman & Croft, 2007; Vihman, 2019), roughly coincide with a child's first referential word use. The practice of speaking, regardless of accuracy, also helps support the development of phonological memory for novel word forms (Keren-Portnoy et al., 2010; Parra et al., 2011; Vihman, 2022).

Such systematic productions of simplified or restructured forms of words to ease the difficulties of challenging words have been found in children learning a wide range of languages. Templatic use may be reflected in high frequency of occurrence coupled with moderate precision, because it indicates use and overuse, and a 'selection' of the chosen shape together with an 'adaptation' of other words which do not have that shape to the chosen shape (Vihman & Velleman, 2000). From the findings of Lou (2020) and the present study, it seems that Mandarin is no exception. Templates are elaborated upon next.

6.2.1 The rise and fall in templatic use

How may templates be represented in Mandarin? Templates are thought to be a support for eventual accurate production of forms, so when will the templates found in the present study start to fade? As templates are constantly being organised and reorganised based on the children's own lexicon, there will come a time where the patterns the children produce are, simply, the accurate adult form. An increase in accuracy of words then corresponds with a decrease in templatic use. For example, Vihman's (2019) reanalysis of Priestly's (1977) child's use of his favoured [CVjVC] template, presented by age of acquisition, shows a distinct rise and then fall in the use of the template. The child selects and adapts many words to the template between weeks 3 and 6 of recording and accurate forms alongside templatic forms around week 4, but from week 10 onwards, most of his productions have become much closer to the adult form and the use of the template has faded.

The findings from this study show that segment and tone templates are independent of one another, just as segment and tone accuracy are independent (based on the experimental findings in Chapter 5). Lou found monosyllabic but not disyllabic segmental templates at the 25wp for some children. In the present study, monosyllables were not analysed, but the disyllabic segmental /t/-/t/ and other /t/-containing consonant sequences found involved substitution patterns for more challenging fricative/affricate sounds and therefore were not considered templatic. Thus findings about disyllabic segmental templates in Mandarin remain inconclusive. Further work could study children at 2.5 to 3 years and beyond to determine whether consonant sequence templates may be found.

However, it seems that there is a different trajectory for tone templates. At the 4wp Lou found templatic use of T1 and T4 and she found no disyllabic tone templates at either word point. The fact that the monosyllabic tone templates found were no longer templatic by the 25wp indicates that children were becoming more accurate in their monosyllabic productions. This may indicate a developmental shift, in that templatic use decreases with increase in vocal practice and accuracy. This is in line with the original idea that templates are mostly found at the one-word stage (Vihman & Croft, 2007; Vihman, 2019). However, even at the 25wp, Lou's children were still inconsistently producing monosyllabic tones and were not yet fully accurate (cf. other studies of inconsistent usage of tones even by children aged 3 years and up by Wong and colleagues, 2005, 2012, 2017). Lou reconciled this contrast by explaining that each of the four tones had a different developmental trajectory. Use of T1 and T4 was far more accurate at the 25wp and these tones were less often used as substitutes for the other tones at that stage; at the same time, T2 and T3 were

increasingly being used and misused, but there was no longer a clear pattern in the monosyllabic tone use.

In this study, in contrast, there were two disyllabic tone templates (and one pattern) beyond the 25wp. The T1-*x* tone templates (specifically, T1-1 and T1-0) found in the present study resemble Mok and Lee's (2018) extension of Vihman's templates, given that the children in the present study were no longer at the one-word stage and these templates were used alongside accurate forms. T1-1 was selected for use in the lexical category of kinship terms and adapted for use in general. T1-0 was overwhelmingly produced by children; it was seemingly their preferred motoric routine in general, which would account for their applying it to word *and* nonword stimuli alike. (T1-4 was also frequently used, but as most uses had high levels of precision, it could not be considered 'templatic'.) Meanwhile, T3-2 is simply a pattern specific to the pronunciations of kinship terms (among other words, as mentioned in section 4.2.3.2.3) in Singapore. Future work could examine the specific prevalence of various tone sequences in other linguistically diverse communities.

The overwhelming use of T1-*x* tone sequences deserves more discussion. Not only is T1 acoustically salient, thus catching children's attention from a young age (cf. Fernald, 1999), but children have also had extended proprioceptive experience with T1. Evidence for this is that prelinguistic children learning Mandarin make considerable use of T1 in babbling (Lou et al., 2018). Lou (2020) has shown that at the 4wp T1 and T4 are used for production ('selected') and overused for the other tones ('adapting' them), again due to the high perceptual salience of these tones and a continuation of the babbling routines children have developed with these tones.

At the 25wp, while they were not templatic, the children in Lou (2020) also favoured tone sequences which start out high and then fall, i.e., T1-4 and T1-0. Even though mothers and children in the present study used a wider range of tone sequences than those in Lou (2020), T1-*x* tone sequences may still have stood out in their everyday speech, thus the high early usage is sustained through to the later lexical stage, represented by the weeks following the 25wp of the children in this study. Evidence for the templatic use of T1-0 can also be observed in children's responses to novel stimuli in the experimental nonword repetition task: T1-0 was the tone sequence of choice.

Chen (J.-Y., 1999, 2000) has suggested that Mandarin lexical tone production is analogous to English stress production, where the stress pattern provides the phonological frame for the word and the segmental information is then inserted into the frame. Examining the

evidence from both production paradigms (cf. Table 6.2), it seems that children target and produce *real words* with frequent segmental sequences that also have a frequent tone sequence (FSFT). In contrast, in direct imitation of *newly encountered stimuli*, within a limited set of stimuli, (FSIT), children produce more words with frequent segmental sequences and infrequent tone sequences (FSIT; some of the T1-*x* sequences children often produce are actually infrequently occurring in the language, but the salience of T1 in S1 seems to take precedence over other frequently occurring tone sequences), but words containing frequent tone sequences (FSFT and ISFT) were still more precisely produced.

The whole-word approach in Mandarin may thus be characterised as follows: a salient and well-practiced tone (T1) provides a tone 'envelope' for segments to fill in (which may then be the foundation for adult studies that find that tone information is processed first relative to syllable information, especially in top-down processing, cf. Liu & Samuel, 2007; Sereno & Lee, 2015; Ye & Connine, 1999). The high acoustic salience allows children to pick up T1 early on, leading them to practice it during babble (Lou et al., 2018). The well-practiced T1-*x* motoric routines thus involve less of a cognitive load, regardless of the fact that the T1-*x* tone sequence may be infrequent in the language (in line with previous findings that children's templates are idiosyncratic and may be independent of ambient language effects, Menn, 2013; Ota, 2013). As a result, attention can be directed to the segmental sequences, which are (50-60%) variegated, so that children may still be fairly precise at matching the variegated word structures in Mandarin. Variegated word structures are the most challenging word structures for children to produce. However, the children in this study were adept at producing variegated responses for variegated targets, as was reported in Lou (2020) and Vihman et al. (2022).

When may consonant sequence templates be found, then, and when will these tone sequence templates begin to fade? Given the lack of 'mastery' of tone found in children even up to 6 years of age (e.g., Wong et al., 2005, 2012, 2017; Xu Rattanasone et al., 2018), tone templates may exist for a while yet, as support for the eventual mastery of the adult forms. Future work with 2.5 to 3-year-olds may provide evidence from the next stage in the production of segmental and tone sequences to see if segmental patterns emerge and tone patterns fade or endure.

6.2.2 Revisit: The (a)symmetry of segmental and tonal information

In the chapter on Mandarin the literature concerning the lexical access of segmental and tonal information and error analyses of segments and tones in Mandarin was discussed, with a consensus that there may be a dissociation in the processing of segment and tone information. Children learning the Chinese languages have rarely been tested with production paradigms outside of attempts to determine the age of mastery of segmental and tonal information. It makes sense to study segmental and tonal information together and ask whether there could also be a dissociation in the way these two pieces of information are processed in production-based tasks in children. The findings from this study support this conclusion: the presence of tone but not segment templates past the 25wp was noted in the naturalistic observations; the fact that frequent segmental sequences or tone sequences had no effect on the accuracy of the other element (tone accuracy and segment accuracy, respectively) might also indicate independence of segmental and tonal elements.

Segments and tones can be said to develop independently and at different speeds (cf. Noiray et al., 2019 on holistic and segmental elements in a language developing together but "at different paces at different times", p. 12). This may apply to Mandarin acquisition too: the developmental trajectory for tone production seems to span a longer time period than that for segment production: for tone, potentially 9 months (cf. Lou et al., 2018, especially for T1) to 6 years (cf. Wong and colleagues, 2005, 2012, 2017), for segments: potentially 2 (cf. Li & To, 2017) to 4 years (cf. Ma et al., 2022, especially for affricates). Tonal information is available to and retained by children more quickly than segmental information (especially in short-term recall), but children take a longer time to master it and become adult-like in its production. In contrast, the window for segments is shorter; segments are produced later than tones but seem to stabilise earlier.

Thus, the shift in focus between segmental and tonal information and between production and perception as children advance lexically may be as follows: T1 is frequent in 9- to 12-month-old children's babbling (Lou et al., 2018), leading to templatic use of T1 and T4 at the 4wp (roughly 13 months old; Lou, 2020). Monosyllabic *segmental* templates are then found at the 25wp (at roughly 18 months), while use of monosyllabic *tone* templates have faded (Lou, 2020). Past the 25wp (from 20 months on, as in the present study), most consonant segments are said to have been acquired (according to studies in Li & To, 2017). However, not all consonant segments were found to have been acquired by the children in this study. Many of the most frequently produced consonant sequences were used to substitute for the more challenging fricative and affricate sounds that children have not yet mastered (cf. Ma et

al., 2022); no disyllabic segmental templates were identified. At this point, though, disyllabic tone templates or patterns may be found (T1-1, T1-0, T3-2). At the stage where disyllabic tone *production* is templatic, *perception* studies are also showing children to be more sensitive to *tone* mispronunciations (Wewalaarachchi et al., 2017). The proprioceptive experience of tone use as well as perception of the corresponding auditory information and 'kinetic feedback' the child receives from hearing themselves speak (McAllister Byun & Tessier, 2016; Vihman, 2022; cf. Icht & Mama, 2015) builds on and reinforces children's phonological memory and sensitivity to tone. The practice that children have gained from the repertoire of consonant segments that children have acquired (given the lack of templates found post-25wp) then allows them to perceive segment mispronunciations more readily from age 4 years onwards (Singh et al., 2015), whereas tone is still being practiced and takes time to master long after segment acquisition (cf. Wong et al., 2005, 2012, 2017; Xu Rattanasone et al., 2018). Thus, there is an interchange between segments and tones and perception and production: tone perception and production begin early but tone production is mastered late, segment perception and production occur later but segment production may still yet be mastered before tone is. As mentioned earlier, future work with 2.5- to 3-year-olds could provide evidence regarding the next stage in the mastery of segment/tone production.

6.3 Implications and future directions

The objective of the study was to systematically investigate the ways in which children learning Mandarin represent the segments and tones in the language they are learning. This study adds to the production literature, which has been studied less, by means of two paradigms. Specifically, both the speech forms children produce in an everyday setting and their responses to unfamiliar stimuli were studied. The thesis integrated two complementary strands of research: the study of short-term retrieval and production practice complemented the study of the influences of long(er)-term knowledge on phonological development. Evidence that segment and tone are independent of one another was presented in both *production* paradigms, in line with much of the previous literature based on adult and child *perception* tasks. Additionally, there was evidence of tonal but not of segmental templatic use in the children's words. Applying the whole-word approach to Mandarin, the salience of T1 provided a tone envelope for the production of segmental sequences.

Perception studies have found a potential three-way asymmetry among consonants, vowels and tones. Future researchers may want to study older children's speech forms to ascertain if there is a corresponding three-way asymmetry in production. Future researchers could also study a larger sample of children to extend the naturalistic observations of this study.

As mentioned in sections 5.3 and 6.1.3, frequency of segmental sequences was not significant until syllable position was included in the analysis of segment accuracy (with more accurate frequent segmental sequences in S2 than S1). Looking at the nonword stimuli, after counterbalancing across the conditions, there were more plosive-initial syllables in **S2** of the nonwords with **frequent segmental sequences**, coupled with more plosive-initial syllables in the **S1** of the nonwords with **infrequent segmental sequences**. Future analyses might benefit from analysing the accuracy by syllable in addition to the whole nonword. This will address the question of whether the driving force behind the responses is simply ease of articulation or familiarity through production experience with a given syllable type (in terms of manner), rather than the frequency of the segmental *sequences* themselves. This will tie in with the proposal in Vihman et al. (2022, 2023) that the syllable, rather than the word, may be the basic unit of segmental representation for Mandarin.

This study contributes to the limited production literature on how children represent segments and tones in Mandarin and the patterns and templates found in their word forms. The linguistically diverse environment provides a rich base on which further in-depth analyses could be conducted. The patterns of phonological development found for these children immersed in this linguistically diverse environment may be of both theoretical and practical interest for educators, linguists and speech language pathologists working with children in other multilingual environments.

Appendices

Appendix A: List of words added to the nonword repetition task

Twenty words that were familiar to 2-year-old children were selected from the Singapore CDI and added to the repetition task to make the task less daunting (Table A1). These words were not analysed.

Table A1: Selected words from Singapore CDI

Word	Gloss
tɕi1tan4	egg
hau3uan2	fun to play with
hu2tiɛ2	butterfly
tʰiɛn1kʰoŋ1	sky
pʰu2tʰao2	grape
kʰuai4tiɛn3	hurry
tsai4tɕiɛn4	goodbye
ɕi3huan1	like
ɕiau3mau1	kitty
hei1sə4	black
tɕiɛn1paŋ3	shoulder
ta2sau3	clean
ta4ɕiaŋ4	elephant
kʰai1ɕin1	happy
tʰai4iaŋ2	sun
paŋ1maŋ2	help
ɕiou1ɕi2	nap
pʰiŋ2kuo3	apple
tɕin1uan3	tonight
kuan4tʰou2	can

References

Afendras, E. A., & Kuo, E. C. Y. (Eds.). (1980). *Language and society in Singapore.* Singapore: Singapore University Press.

Altvater-Mackensen, N. & Fikkert, P. (2015). A Cross-Linguistic Perspective on the Acquisition of Manner of Articulation Contrasts in the Productions of Dutch and German Children. *Language Acquisition*, *22*(1), 2-39. https://doi.org/10.1080/10489223.2014.892945

Amayreh, M. M., & Dyson, A. T. (1998). The acquisition of Arabic consonants. *Journal of Speech, Language, and Hearing Research*, *41*(3), 642-653. https://doi.org/10.1044/jslhr.4103.642

Ambridge, B., Kidd, E., Rowland, C., & Theakston, A. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239-273. https://doi.org/10.1017/S030500091400049X

Ansaldo, U. (2010). Contact and Asian Varieties of English. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 498-417). John Wiley & Sons.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*(5), 970-983. https://doi.org/10.1044/1092-4388(2006/070)

Baayen, R. H., Piepenbrock, R., & Gulikersm, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania. [cited in Ke (2006)]

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173. https://doi.org/10.1037/0033-295x.105.1.158

Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015) Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beckman, M. E., & Edwards, J. (2000). Lexical frequency effects on young children's imitative productions. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 207-217). Cambridge University Press.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin*, *130*(6), 858-886. https://doi.org/10.1037/0033-2909.130.6.858

Boersma, P., & Weenik, D. (2012). Praat: Doing phonetics by computer [Computer program]. Version 5.1. Retrieved from http://www.fon.hum.uva.nl/praat/

Bowey, J. A. (1997). What *Does* Nonword Repetition Measure? A Reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology*, *67*(2), 295-301. https://doi.org/10.1006/jecp.1997.2408

Brehm, L. & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*, 104334.

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology*, *108*(4), 693-712. https://doi.org/10.1016/j.jecp.2010.07.008

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (Language Learning and Language Teaching series) (pp. 259-280). Amsterdam: John Benjamins.

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(3), 455-474. https://doi.org/10.1017/S0142716407070257

Bybee, J. L. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216-236). New York: Routledge

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, *10*(2), 159-199. https://doi.org/10.1016/0885-2014(95)90008-X

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, *8*(1), 77-111. https://doi.org/10.1075/ni.8.1.05cha

Chao, Y. R. (1930). A system of 'tone-letters'. *Le Maître Phonétique, 45*, 24-27.

Chao, Y. R. (1951/1973). The Cantian ideolect: An analysis of the Chinese spoken by a twenty-eight-months-old child. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart and Winston, INC.

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

Chen, C., Wang, M., Shu, H., Wu, H., & Li, C. C. (2010). Development of Tone Sensitivity in Young Chinese Children. In *The Proceedings of the Poster Presentations at Speech Prosody*, Chicago, IL.

Chen, C. Y. 陈重瑜 (1983). 新加坡华语－语法与词汇特征 *Xinjiapo Huayu – Yufa yu Cihui Tezheng* (Features of Singapore Mandarin: Syntactic and Lexical). Centre for Chinese Language, National University of Singapore.

Chen, C. Y. 陈重瑜 (1984). Neutral tone in Mandarin: phonotactic description and the issue of the norm. *Journal of Chinese Linguistics*, *12*(2), 209-332. https://www.jstor.org/stable/23767004

Chen, C. Y. 陈重瑜 (1986). 新加坡华语语法特征 *Xinjiapo Huayu Yufa Tezheng* (Syntactic features of Singapore Mandarin). *Yuyan Yanjiu (Studies in Language and Linguistics)*, 1, 138-152.

Chen, J.-Y. (1993). A corpus of speech errors in Mandarin Chinese and their classification. *The World of Chinese Language*, *69*, 26-41. [cited in Wan & Jaeger, 1998]

Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics*, *20*(2), 289-301. https://doi.org/10.1017/S0142716499002064

Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient*, *43*(1), 15-26.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*(4), 751-781. https://doi.org/10.1006/jmla.2001.2825

Chen, L.-M. & Kent, R. D. (2005). Consonant-vowel co-occurrence patterns in Mandarin-learning infants. *Journal of Child Language*, *32*(3), 507-534. https://doi.org/10.1017/S0305000905006896

Chen, L.-M. & Kent, R. D. (2009). Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language*, *36*(1), 73-84. https://doi.org/10.1017/S0305000908008878

Chen, P. (1999). *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press.

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. https://doi.org/10.1159/000091406

Cheng, C.-M. (1982). Analysis of present-day Mandarin (abstract). *Journal of Chinese Linguistics*, *10*(2), 281-358. https://www.jstor.org/stable/23767014

Chew, C. H. 周清海 (2007). 论全球化环境下华语的规范问题 *Lun quanqiu yujing xia huayu de guifan wenti* (Standardization of the Chinese language: A global language). 语言教学与研究 *Yuyan Jiaoxue Yu Yanjiu (Language Teaching and Research)*, *4*, 91-96.

Ching, N. C. (1971). A minor problem in tone sandhi. *Journal of the Chinese Language Teachers' Association*, *6*(2), 41-57. [cited in Chen, (1984)]

Cho, S. J. (2008). The Phonological development of Taiwan preschoolers from age three to six. Unpublished Master's thesis. National Taipei University of Nursing and Health Sciences, Taiwan. [cited in Li & To (2017)]

Choo, R. Q., Vihman, M. & Keren-Portnoy, T. (2019). The acquisition of lexical tones by Mandarin-English bilinguals: A longitudinal study. Talk presented at the 12th International Symposium of Bilingualism, Edmonton, Alberta, Canada, 23-28 June 2019.

Chua, C. L. (2003). *The emergence of Singapore Mandarin: A case study of language contact.* PhD thesis, University of Wisconsin-Madison.

Clements, G. N. (1991). Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory*, *5*, 77-123.

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

Clumeck, H. V. (1980). The acquisition of tone. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology Volume 1: Production* (pp. 257-275). New York: Academic Press.

Coady, J. A. & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language and Communication Disorders*, *43*(1), 1-40. https://doi.org/10.1080/13682820601116485

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the Beginnings of Spoken Words Have a Special Status in Auditory Word Recognition? *Journal of Memory and Language*, *32*(2), 193-210. https://doi.org/10.1006/jmla.1993.1011

Crelin, E. S. (1987). *The Human Vocal Tract: Anatomy, Function, Development, and Evolution*. New York, NY: Vantage Press.

Crystal, D. (2008). *Dictionary of Linguistics and Phonetics (6th Ed.)*. Oxford, UK: Blackwell Publishing

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*(3-4), 133-142. https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., Mehler, J. Norris, D. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*(4), 385-400. https://doi.org/10.1016/0749-596X(86)90033-1

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research, 33*(1), 16-27. https://doi.org/10.1044/jshr.3301.16

Davis, B. L. & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*(6), 1199-1211. https://doi.org/10.1044/jshr.3806.1199

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of Serial Complexity in Speech Production: A Comparison of Phonetic and Phonological Approaches to First Word Production. *Phonetica, 59*(2-3), 75-107. https://doi.org/10.1159/000066065

DeFrancis, J. (1984). *The Chinese Language: Fact and Fantasy*. Honolulu: University of Hawaii Press.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: a study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(6), 1355-1367. https://doi.org/10.1037//0278-7393.26.6.1355

Demuth, K. & Johnson, M. (2003). Truncation to Subminimal Words in Early French. *Canadian Journal of Linguistics, 48*(3/4), 211-241. https://doi.org/10.1017/S0008413100000657

Demuth, K., & Johnson, M. (2020). Exemplar-based learning probably requires learning abstractions: A commentary on Ambridge (2020). *First Language, 40*(5–6), 573–575. https://doi.org/10.1177/0142723720903609

Deng, L. & Dang, J. (2007). Speech analysis: The Production-Perception perspective. In H.-Z. Li & C.-H. Lee (Eds.), *Advances in Chinese spoken language processing* (pp. 3-32). New Jersey: World Scientific.

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: the Tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69-92. https://www.jstor.org/stable/26538063

DePaolis, R., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior & Development*, *34*(4), 590-601. https://doi.org/doi:10.1016/j.infbeh.2011.06.005

DePaolis, R., Vihman, M. M., & Nakai, S. (2013). The influence of babbling patterns on the processing of speech. *Infant Behavior and Development*, *36*(4), 642-649. https://doi.org/10.1016/j.infbeh.2013.06.007

Department of Statistics Singapore. (2020). *Chinese Resident Population by Age Group, Dialect Group and Sex, 2015*. Retrieved July 31, 2021 from the website https://data.gov.sg/dataset/chinese-resident-population-by-age-group-dialect-group-and-sex-2015

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense word repetition task. *Journal of Speech and Hearing Research*, *36*(5), 1051-1054. https://doi.org/10.1044/jshr.3605.1051

Dollaghan, C., Biber, M., & Campbell, T. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, *16*(2), 211-222. https://doi.org/10.1017/S0142716400007098

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Edwards, J., Beckman, M. E. & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research*, *47*(2), 421-436. https://doi.org/10.1044/1092-4388(2004/034)

Edwards, J., Munson, B. & Beckman, M. E. (2011). Lexicon-phonology relationships and dynamics of early language development – A commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language*, *38*(1), 35-40. https://doi.org/10.1017/S0305000910000450

Eimas, P. D. (1978). Developmental aspects of speech perception. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (Vol. 8). Berlin: Springer-Verlag

ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865-878. https://doi.org/10.1044/jslhr.4304.865

Endress, A., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, *105*(3), 577-614. https://doi.org/10.1016/j.cognition.2006.12.014

Erbaugh, M, S. (1992). The acquisition of Mandarin. In D. I. Slobin (Ed.), *The Cross-linguistic Study of Language Acquisition, Volume 3* (pp. 373-455). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Farwell, C. B. (1976). Some strategies in the early production of fricatives. *Papers and Reports in Child Language Development*, *12*, 97-104. Stanford University Linguistics Department.

Feng, C., Yue, Y. & Zhang, Q. (2019). Syllables are Retrieved before Segments in the Spoken Production of Mandarin Chinese: An ERP Study. *Scientific Reports*, *9*, 11773. https://doi.org/10.1038/s41598-019-48033-3

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J. (1993). The MacArthur-Bates Communicative Development Inventories: User's guide and technical manual. San Diego, CA: Singular Publishing Group

Ferguson, C. A. & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, *51*(2), 419-439. Reprinted in Vihman & Keren-Portnoy, 2013.

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*(6), 1497-1510. https://doi.org/10.2307/1130938

Fromkin, V. A. (Ed.) (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Gandour, J. (1977). Counterfeit tones in the speech of southern Thai bidialectals. *Lingua*, *41*(2), 125-143. https://doi.org/10.1016/0024-3841(77)90054-7

Gandour, J. T., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A Crosslinguistic PET Study of Tone Perception. *Journal of Cognitive Neuroscience*, *12*(1), 207-222. https://doi.org/10.1162/089892900561841

Gao, J., Shi, R., & Li, A. (2011). Lexical tone representations of Mandarin-speaking toddlers. Poster presented at the International Congress of Phonetic Sciences, Hong Kong.

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513-543. https://doi.org/10.1017/S0142716406060383

Gathercole, S. E., & Adams, A.-M. (1993). Phonological working memory in very young children. *Developmental Psychology*, *29*(4), 770-778. https://doi.org/10.1037/0012-1649.29.4.770

Gathercole, S. E. & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336-360. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E. & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.

Gathercole, S. E., Frankish, C., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84-95. https://doi.org/10.1037/0278-7393.25.1.84

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, *33*(6), 966-979. https://doi.org/10.1037//0012-1649.33.6.966

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, *12*(3), 349-367. https://doi.org/10.1017/S0142716400009267

Gildersleeve-Neumann, C., Davis, B. & MacNeilage, P. (2000). Contingencies governing the production of fricatives, affricates, and liquids in babbling. *Applied Psycholinguistics*, *21*(3), 341–363. https://doi.org/10.1017/S0142716400003039

Gordon, R. G., Jr. (Ed.). (2005). *Ethnologue: Languages of the World* (15th ed.). Dallas, Texas: SIL International.

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual Reorganization of Lexical Tones: Effects of Age and Experimental Procedure. *Frontiers in Psychology*, *9*, 477. https://doi.org/10.3389/fpsyg.2018.00477

Guo, S. 郭绍虞 (1938). 中国语词之弹性作用 *Zhongguo yuci zhi tanxing zuoyong* (The function of elastic word length in Chinese). 燕京学报 *Yanjing Xuebao (Yanjing Journal)*, *24*, 1-34.

Guo, X. 郭熙 (2007). 以 '国、中、汉、华、唐' 为上字的词与社会认同建构 *Yi 'Guo, Zhong, Han, Hua, Tang' wei shangzi de ci yu shehui rentong jiangou* (The words beginning with Guo, Zhong, Han, Hua, Tang and the Construction of Social Identification). 语言教学与研究 *Yuyan Jiaoxue yu Yanjiu (Language Teaching and Research)*, *4*, 83-90.

Ho, C. S.-H. & Lai, D. N.-C. (1999). Naming-speed deficits and phonological memory deficits in Chinese developmental dyslexia. *Learning and Individual Differences*, *11*(2), 173-186. https://doi.org/10.1016/S1041-6080(00)80004-7

Hockett, C. F. (1947). Peiping phonology. *Journal of the American Oriental Society*, *67*(4), 253-267. https://doi.org/10.2307/596062

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, *35*(4), 903-916. https://doi.org/10.1017/S0305000908008751

Hsu, J. H. (1996). *A study of the stages of development and acquisition of Mandarin Chinese by children in Taiwan*. Taipei, Taiwan: Crane Publishing. (cited in Li & To, 2017)

Huff, C. L. (2017). *A Study of Mandarin Homophony*. Unpublished doctoral thesis, University of Cambridge, UK.

Icht, M. & Mama, Y. (2015). The production effect in memory: A prominent mnemonic in children. *Journal of Child Language*, *42*(5), 1102-1124. https://doi.org/10.1017/S0305000914000713

Jakobson, R. (1941/1968) Child Language, Aphasia and Phonological Universals, The Hague: Mouton.

Jeng, J. Y. (2011). The phonological processes of syllable-initial consonants spoken by the preschool children of Mandarin Chinese (in Chinese). *Journal of Special Education*, *34*, 135-169. [cited in Li & To, 2017]

Jeng, J. Y. (2013). The development of speech sound assessment battery for Mandarin speaking children. Retrieved from http://giast.nknu.edu.tw/UploadFile/TeaFiles/giast_t_103205415.pdf

Jin, W. (2011). A statistical argument for the homophony avoidance approach to the disyllabification of Chinese. In Z. Jing-Schmidt (Ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics* (NACCL-23), *1*, 35-50. University of Oregon, Eugene.

Jones, D. M., & Macken, B. (2018). In the beginning was the deed: Verbal Short-Term Memory as Object-Oriented Action. *Current Directions in Psychological Science*, *27*(5), 351-356. https://doi.org/10.1177/09637214187657

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The Phonological Store of Working Memory: Is It Phonological and Is It a Store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 656-674. https://doi.org/10.1037/0278-7393.30.3.656

Jusczyk, P. W. (1986). Toward a model of the development of speech perception. In J. S. Perkell and D. H. Klatt (Eds.) *Invariance and Variability in Speech Processes* (pp. 1-35). Hillsdale, New Jersey: Erlbaum

Karlgren, B. (1949). *The Chinese language: an essay on its nature and history*. New York: The Ronald Press Co.

Katzner, K. (1977). *The Languages of the World*. London: Routledge & Kegan Paul.

Ke, J. (2006). A cross-linguistic quantitative study of homophony. *Journal of Quantitative Linguistics*, *13*(1), 129-159. https://doi.org/10.1080/09296170500500850

Kehoe, M. (2015). Lexical-phonological interactions in bilingual children. *First Language*, *35*(2), 93-125. https://doi.org/10.1177/0142723715574398

Kent, R. D., & Vorperian, H. K. (1995). Development of the craniofacial-oral-laryngeal anatomy: a review. *Journal of Medical Speech Language Pathology*, *3*, 145-190.

Keren-Portnoy, T., Vihman, M. M., DePaolis, R., Whitaker, C., & Williams, N. A. (2010). The role of vocal practice in constructing phonological working memory. *Journal of Speech, Language, and Hearing Research*, *53*(5), 1280-1293. https://doi.org/10.1044/1092-4388(2009/09-0003)

Khattab, G., & Al-Tamimi, J. (2013). Influence of geminate structure on early Arabic templatic patterns. In M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-word Approaches and Cross-linguistic Evidence* (pp. 374-414). Cambridge University Press. https://doi.org/10.1017/CBO9780511980503.018

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A Cross-Linguistic PET Study of Tone Perception in Mandarin Chinese and English Speakers. *NeuroImage*, *13(*4*)*, 646-653. https://doi.org/10.1006/nimg.2000.0738

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech*, *32*(3), 249-278. https://doi.org/10.1177/002383098903200304

Kučera, H. & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K. & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science*, *255*(5044), 606-608. https://doi.org/10.1126/science.1736364

Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, *5*(2), 13-24.

Lee, L. (2010). The Tonal System of Singapore Mandarin. In Clemens, L.E. & C.-M. L. Liu, (Eds.), *Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18) (Vol. 1)*, pp. 345-362. Harvard University, Cambridge, MA.

Lee, C. L. (2011). *Politeness in Singapore*. In D.Z. Kádár & Sara Mills (Eds.), *Politeness in East Asia* (pp. 236-251).

Lee, C.-Y. (2007). Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech*, *50*(1), 101-123. https://doi.org/10.1177/00238309070500010501

Lee, E. F. F. (2001). Profile of the Singapore Chinese Dialect Groups. *Statistics Singapore Newsletter* (p. 2-6). Singapore Department of Statistics.

Lee, M. L. C. & Rickard Liow, S. J. (2015). *Cognitive-Linguistic Assessment Profile – Bilingual Language Assessment Battery for Children*. Unpublished measure. Yong Loo Lin School of Medicine, Division of Graduate Medical Studies, National University of Singapore.

Leonard, L. B., Schwartz, R. G., Morris, B., & Chapman, K. (1981). Factors influencing early lexical acquisition: Lexical orientation and phonological composition. *Child Development*, *52*(3), 882-887. https://doi.org/10.2307/1129090

Levelt, W. J. M. (1993). Lexical Access in Speech Production. In E. Reuland & W. Abraham (Eds.), *Knowledge and Language, Volume I: From Orwell's Problem to Plato's Problem* (pp. 241-251). The Netherlands: Kluwer Academic Publishers.

Levelt, C. (1994). *On the acquisition of place*. Doctoral thesis, Rijksuniversiteit Leiden.

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, *22*(1), 1-75. https://doi.org/10.1017/S0140525X99001776

Li, F., & Munson, B. (2016). The development of voiceless sibilant fricatives in Putonghua-speaking children. *Journal of Speech and Language Hearing Research*, *59*(4), 699-712. https://doi.org/10.1044/2016_JSLHR-S-14-0142

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4(2)*, 185-199. https://doi.org/10.1017/S0305000900001598

Li, C. N. & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.

Li, X. X., & To, C. (2017). A Review of Phonological Development of Mandarin-Speaking Children. *American Journal of Speech-Language Pathology*, *26*(4), 1262-1278. https://doi.org/10.1044/2017_AJSLP-16-0061

Lim, L. (2007). Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes*, *26*(4), 446-473. https://doi.org/10.1111/j.1467-971X.2007.00522.x

Lin, H. (1990). Towards an analysis of Mandarin reduplication. *Working Papers of the Linguistics Circle*, 129-141.

Lin, C., Wang, M., & Shu, H. (2013). The processing of lexical tones by young Chinese children. *Journal of Child Language*, *40*(4), 885-899. https://doi.org/10.1017/S0305000912000281

Lin, T. & Wang, L. (1992). *Yuyinxue Jiaocheng (A course of phonetics)*. Beijing: Peking University

Lin, M. & Yan, J. (1980). 北京话轻声的声学性质 *Beijinghua qingsheng de shengxue xingzhi* (The acoustic nature of the Mandarin Neutral Tone). 方言 *Fangyan (Dialect)*, *3*, 166-178. [cited in Chen & Xu, 2006]

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385-394. https://doi.org/10.1016/j.cognition.2014.06.004

Liu, L., Ong, J. H., Alba, T. & Paola, E. (2018). One Way or Another: Evidence for Perceptual Asymmetry in Pre-attentive Learning of Non-native Contrasts. *Frontiers in Psychology*, *9*, 162. https://doi.org/10.3389/fpsyg.2018.00162

Liu, S. & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, *22(4)*, 566-594. https://doi.org/10.1080/01690960600989600

Lleó, C. (2002). The role of markedness in the acquisition of complex prosodic structures by German-Spanish bilinguals. *International Journal of Bilingualism*, *6*(3), 233-237. https://doi.org/10.1177/13670069020060030501

Lock, G. (1989). Aspects of variation and change in the Mandarin Chinese spoken in Singapore. *Australian Journal of Linguistics*, *9*(2), 277-294. https://doi.org/10.1080/07268608908599423

Locke, J. (1983). *Phonological Acquisition and Change*. New York: Academic Press.

López Assef, B. L., Desmeules-Trudel, F., Bernard, A. & Zamuner, T. S. (2021). A shift in the direction of the production effect in children aged 2-6 years. *Child Development*, *92*(6), 2447-2464. https://doi.org/10.1111/cdev.13618

Lou, S., Vihman, M. M., & Keren-Portnoy, T. (2018). *Ambient-Language Effects in Babbling: A study of Mandarin- and English-learning infants.* Posted presented at the Child Language Symposium 2018, Reading, UK, 25-26 June 2018

Lou, S. (2020). Early phonological development in Mandarin: An analysis of word structures, segments and tones from babbling through the single-word period. Unpublished doctoral thesis, University of York, UK.

Lü, S. (1963). 现代汉语单双音节问题初探 *Xiandai Hanyu dan shuang yinjie wenti chu tan (A preliminary study of the problem of monosyllabism and disyllabism in modern Chinese). Zhongguo Yuwen (Chinese Language),* 1, 11-23.

Luo, C.-C. (1992). *Kuo Yü hsüeh* (Studies of Mandarin Chinese). Taipei: Wu-Nan

Ma, J., Wu, Y., Zhu, J. & Chen, X. (2022) The Phonological Development of Mandarin Voiceless Affricates in Three- to Five-Year-Old Children. *Frontiers in Psychology*, *13*, 809722. https://doi.org/10.3389/fpsyg.2022.809722

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017). Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition*, *159*, 139-155. https://doi.org/10.1016/j.cognition.2016.11.011

Macken, M. A. (1979/2013). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. In M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-word Approaches and Cross-linguistic Evidence* (pp. 133-167). Cambridge: Cambridge University Press.

Macken, M. A. (1980). Aspects of the acquisition of stop systems: A cross-linguistic perspective. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology Volume 1: Production* (pp. 143-168). New York: Academic Press.

MacNeilage, P. F. & Davis, B. L. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and Performance XIII: Motor Representation and Control* (pp. 453-476). New Jersey: Lawrence Erlbaum Associates.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk, 3rd ed*. Mahwah, NJ: Lawrence Erlbaum Associates

Maddieson, I. (1978). Universals of tone. In J. H Greenberg, et al. (Eds.), *Universals of Human Language, Volume 2: Phonology* (pp. 335-363). Stanford University Press.

Maddieson, I. (2011). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chapter 13), available online at http://wals.info/chapter/13. Munich, Germany: Max Planck Digital Library

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-71. https://doi.org/10.1016/0010-0277(80)90015-3

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: evidence for perceptual reorganization. *Infancy*, *10*(3), 241-265. https://doi.org/10.1207/s15327078in1003_3

Matthews, S., & Yip, V. (2011). *Cantonese: a comprehensive grammar* (2nd Ed.). New York: Routledge.

McAllister Byun, T. (2012). Positional velar fronting: An updated articulatory account. *Journal of Child Language*, *39*(5), 1043-1076. https://doi.org/10.1017/S0305000911000468

McAllister Byun, T. & Tessier, A.-M. (2016). Motor Influences on Grammar in an Emergentist Model of Phonology. *Language and Linguistics Compass*, *10*(9), 431-452. https://doi.org/10.1111/lnc3.12205

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. https://doi.org/10.1016/0010-0285(86)90015-0

Menn, L. (1976). Pattern, control and contrast in beginning speech: a case study in the acquisition of word form and function. Unpublished PhD thesis, University of Illinois. Circulated by Indiana University Linguistics Club.

Menn, L. (1983/2013). Development of articulatory, phonetic, and phonological capabilities In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches, cross-linguistic evidence* (pp. 168-214). Cambridge, UK: Cambridge University Press

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Lawrence Erlbaum Associates Publishers.

Mok, P. P. K., Fung, H. S. H., & Li, V. G. (2019). Assessing the Link Between Perception and Production in Cantonese Tone Acquisition. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1243-1257. https://doi.org/10.1044/2018_JSLHR-S-17-0430

Mok, P. P. K. & Lee, A. (2018). The acquisition of lexical tones by Cantonese-English bilingual children. *Journal of Child Language*, *45*(6), 1357-1376. https://doi.org/10.1017/S0305000918000260

Mok, P., Li, V. G., & Fung, H. (2020). Development of Phonetic Contrasts in Cantonese Tone Acquisition. *Journal of Speech, Language, and Hearing Research*, *63*(1), 95-108. https://doi.org/10.1044/2019_JSLHR-19-00152

Mowrer, D. E. & Burger, S. (1991). A comparative analysis of phonological acquisition of consonants in the speech of 2½-6-year-old Xhosa- and English-speaking children. *Clinical Linguistics & Phonetics*, *5*(2), 139-164. https://doi.org/10.3109/02699209108985510

Munson, B., Edwards, J., & Beckman, M. E. (2011). Phonological representations in language acquisition: Climbing the ladder of abstraction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 288–209). Oxford: Oxford University Press.

Ng, P. C. L. (2010). Reflections on Singapore's Speak Mandarin Campaign. *Proceedings of the NEAR Language Education Conference 2010*. http://nearconference.weebly.com/uploads/1/2/7/1/12718010/near-2010-05.pdf

Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S. & Hintermeier, L. (2019). Spoken Language Development and the Challenge of Skill Integration. *Frontiers in Psychology*, *10*, 2777. https://doi.org/10.3389/fpsyg.2019.02777

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

O'Seaghdha, P. G., & Chen, J.-Y. (2009). Toward a language-general account of word production: The proximate units principle. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 68-73). Austin, TX: Cognitive Science Society.

O'Seaghdha P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115(2)*, 282-302. https://doi.org/10.1016/j.cognition.2010.01.001

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2013). Close but not proximate: the significance of phonological segments in speaking depends on their functional engagement. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), E3. https://doi.org/10.1073/pnas.1217032110

Ohala, J. J. (1978). Production of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5-32). New York: Academic Press

Oliveira-Guimarães, D. (2013). Beyond early words: word template development in Brazilian Portuguese. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches, cross-linguistic evidence* (pp. 291-316). Cambridge, UK: Cambridge University Press.

Ota, M. (2013). Lexical frequency effects on phonological development: the case of word production in Japanese. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches, cross-linguistic evidence* (pp. 415-438). Cambridge, UK: Cambridge University Press

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development*, *12*(4), 380-397. https://doi.org/10.1080/15475441.2016.1165100

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of Child Language*, *45*(1), 204-218. https://doi.org/10.1017/S0305000916000660

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, *32*(2), 258-278. https://doi.org/10.1006/jmla.1993.1014

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology*, *108*(1), 113-125. https://doi.org/10.1016/j.jecp.2010.07.011

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2-3), 115-154. https://doi.org/10.1177/00238309030460020501

Priestly, T. M. S. (1977). One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language, 4*(1), 45-66. https://doi.org/10.1017/S0305000900000477

Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition in English and K'iché, in K. E. Nelson and A. van Kleeck (Eds.), Children's Language (pp. 175-190). Hove: Psychology Press https://doi.org/10.4324/9781315792668-8

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109(35)*, 14265-14270. https://doi.org/10.1073/pnas.1200632109

Quam, C., & Swingley, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, *62(2)*, 135-150. https://doi.org/10.1016/j.jml.2009.09.003

R Development Core Team. (2018). *R: A language and environment for statistical computing,* version 3.5.0.4. Vienna, Austria. ISBN: 3-900051-07-0. <http://www.R-project.org>.

Rappa, A. & Wee, L. (2007). *Language policy and modernity in Southeast Asia*. New York: Springer.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223-234. https://doi.org/10.1044/1092-4388(2004/019)

Sagart, L., Halle, P., de Boysson-Bardies, B., and Arabia-Guidet, C. (1986). Tone production in Modern Standard Chinese: An electromyographic investigation. *Cahiers Linguistique Asie-Orientale*, *15*(2), 205-221. https://doi.org/10.1163/19606028-90000434

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders*, *45*(1), 357-377. https://doi.org/10.1044/jshd.3701.55

Savinainen-Makkonen, T. (2007). Geminate template: A model for first Finnish words. *First Language*, *27*(4), 347-359. https://doi.org/10.1177/0142723707081728

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Schwartz, R., & Leonard, L. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*(2), 319-336. https://doi.org/10.1017/S0305000900004748

Sebastian-Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31*(1), 18-32. https://doi.org/10.1016/0749-596X(92)90003-G

Sereno, J. A., & Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. *Language and Speech*, *58(2)*, 131-151. https://doi.org/10.1177/0023830914522956

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology*, *45(1)*, 21-50. https://doi.org/10.1080/14640749208401314

Service, E., & Kohonen, V. (1995). Is the relationship between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16(2)*, 155-172. https://doi.org/10.1017/S0142716400007062

Shang, G. W. & Zhao, S. H. (2013). 华语规范化的标准与路向: 以新加坡华语为例 *Huayu Guifanhua De Biaozhun Yu Luxiang: Yi Xinjiapo Huayu Weili (The standard and direction of Huayu codification: A case study of Singapore Mandarin)*. *Yuyan Jiaoxue Yu Yanjiu (Language Teaching and Research)*, *3*, 82–90.

Shang, G. W. & Zhao, S. H. (2017). What standard and whose standard: Issues in the development of Chinese Proficiency Descriptors in Singapore. In D. Zhang & C.-H. Lin. (Eds.) Chinese as a Second Language Assessment (pp. 159-182). Springer Nature Singapore.

Shattuck-Hufnagel, S. (1987). The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17-51). Hillsdale, NJ: Erlbaum.

Shih, C. (1988). *Tone and Intonation in Mandarin*. In N. Clements (Ed.), *Working Papers of the Cornell Phonetics Laboratory, CLC Publications*, 3, 83-109. [cited in Lee, 2010]

Singh, L., Cheng, Q., Tan, S. H., Tan, A., & Low, Y. L. (2022). Language acquisition in a multilingual society: English vocabulary norms and predictors in Singaporean children. *Child Development*, *93*(1), 288-305. https://doi.org/10.1111/cdev.13676

Singh, L., & Fu, C. S. L. (2016). A new view of language development: the acquisition of lexical tone. *Child Development*, *87*(3), 834-854. https://doi.org/10.1111/cdev.12512

Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, *142*, 1-11. https://doi.org/10.1016/j.cognition.2015.05.010

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*(2), 157-178. https://doi.org/10.1080/15475440801922131

Smalle, E., Muylle, M., Szmalec, A., & Duyck, W. (2017). The different time course of phonotactic constraint learning in children and adults: Evidence from speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1821-1827. https://doi.org/10.1037/xlm0000405

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(3), 369-373. https://doi.org/10.1017/S0142716400009279

So, L. K. H., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, *22*(3), 473-495. https://doi.org/10.1017/S0305000900009922

So, L. K. H., & Zhou, J. (2000). *Putonghua Segmental Phonology Test*. China: Nanjing Normal University. [cited in Li & To, 2017]

Sosa, A. V. (2008). Lexical effects in typical phonological acquisition. Unpublished doctoral thesis, University of Washington, Seattle.

Sosa, A. V. & Stoel-Gammon, C. (2012). Lexical and Phonological Effects in Early Word Production. *Journal of Speech, Language, and Hearing Research*, *55*(2), 596-608. https://doi.org/10.1044/1092-4388(2011/10-0113)

Stoel-Gammon, C. & Stemberger, J. (1994). Consonant Harmony and Underspecification in Child Phonology. In M. Yavas (Ed.), *First and Second Language Phonology* (pp. 63-80). San Diego: Singular Publishing Group, Inc.

Stoel-Gammon C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 1-34. https://doi.org/10.1017/S0305000910000425

Stimson, H. M. (1967). Stress in Peking phonotactics. *Monumenta Serica, Journal of Oriental Studies*, 26, 202-212. https://www.jstor.org/stable/40725849

Stokes, S. F., Wong, A. M-Y., Fletcher, P., & Leonard, L. B. (2006). Non-word repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, *49*(2), 219-236. https://doi.org/10.1044/1092-4388(2006/019)

Storkel, H. L. (2005). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language and Hearing Research*, *47*(5), 1194-1212. https://doi.org/10.1044/1092-4388(2004/088)

Storkel, H. L. (2006). Do children still pick and choose? The relationship between phonological knowledge and lexical acquisition beyond 50 words. *Clinical Linguistics and Phonetics*, *20*(7-8), 523-529. https://doi.org/10.1080/02699200500266349

Storkel, H. L. & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*(2), 497-506. https://doi.org/10.3758/brm.42.2.497

Stroud, C. & Wee, L. (2007) Consuming identities: Language planning and policy in Singaporean late modernity. *Language Policy*, *6*(2), 253-279. https://doi.org/10.1007/s10993-007-9047-6

Stroud, C. & Wee, L. (2010). Language policy and planning in Singaporean late modernity. In L. Lim, A. Pakir & L. Wee (Eds.), *English in Singapore: Modernity and Management* (*pp*. 181-204). Hong Kong University Press.

Styles, S., J., Woon, F. T., Chua, Y. H. V., Fong, S., Wong, J. Y., & Le, T. A. (2022). *Language resources, language choices, and translanguaging in parent/child interactions*. Poster presented at the third biennial Workshop on Infant Language Development (WILD), 9-11 June 2022, San Sebastian, Spain.

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019a). Acquisition of weak syllables in tonal languages: acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, *46*(1), 24-50. https://doi.org/10.1017/S0305000918000296

Tang, P., Yuen, I., Rattanasone, N. X., Gao, L., & Demuth, K. (2019b). The acquisition of phonological alternations: the case of the Mandarin tone sandhi process. *Applied Psycholinguistics*, *40*(6), 1495-1526. https://doi.org/10.1017/S0142716419000353

Thelen, E. (1981). Rhythmical behavior in infancy. *Developmental Psychology*, *17*(3), 237-257. https://doi.org/10.1037/0012-1649.17.3.237

To, C. K. S., Cheung, P. S. P., & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels and tones. *Journal of Speech, Language, and Hearing Research*, *56*(1), 103-122. https://doi.org/10.1044/1092-4388(2012/11-0080)

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*(5), 689-708. https://doi.org/10.1080/01690960701728261

Tsao, F.-M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in Psychology*, *8*, 558. https://doi.org/10.3389/fpsyg.2017.00558

Tseng, C.-H. (1998). Syllable as the segmentation unit in perceiving spoken Chinese. *The Journal of the Acoustical Society of America*, *103*(5), 2981. https://doi.org/10.1121/1.422437

Tyler, A. & Edwards, M. L. (1993). Lexical acquisition and the acquisition of initial voiceless stops. *Journal of Child Language*, *20(2)*, 253-273. https://doi.org/10.1017/S0305000900008278

Vihman, M. M. (1978). Consonant harmony: Its scope and function in child language. *Phonology*, *2*, 281-334.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*(1-2), 61-82. https://doi.org/10.1016/S0095-4470(19)31321-X

Vihman, M. M. (2016). Word structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*, *19*(1), 69-88. https://doi.org/10.1017/S1366728914000790

Vihman, M. M. (2017). Learning words and learning sounds. *British Journal of Psychology*, *108*(1), 1-27. https://doi.org/10.1111/bjop.12207

Vihman, M. M. (2019). *Phonological Templates in Development*. Oxford: Oxford University Press.

Vihman, M. M. (2022). Developmental origins of phonological memory. *Psychological Review, 129*(6), 1495-1508. https://doi.org/10.1037/rev0000354

Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics*, *45*(4), 683-725. https://doi.org/10.1515/LING.2007.021

Vihman, M., Ferguson, C., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, *7*(1), 3-40. https://doi.org/10.1017/S0142716400007165

Vihman, M. M., & Keren-Portnoy, T. (Eds.) (2013). *The emergence of phonology: Whole-word approaches, cross-linguistic evidence*. Cambridge, UK: Cambridge University Press

Vihman, M., & Majorano, M. (2017). The role of geminates in infants' early word production and word-form recognition. *Journal of Child Language*, *44*(1), 158-184. https://doi.org/10.1017/S0305000915000793

Vihman, M. M., & McCune, L. (1994). When is a word a word? *Journal of Child Language*, *21*(3), 517–542. https://doi.org/10.1017/S0305000900009442

Vihman, M. M. & Miller, R. (1988). Words and babble at the threshold of lexical acquisition. In M. D. Smith & J. L. Locke (Eds.), *The Emergent Lexicon: The Child's Development of a Linguistic Vocabulary* (pp. 151-183). New York: Academic Press.

Vihman, M. M. & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, *57*(2-4), 255-266. https://doi.org/10.1159/000028478

Vihman, M. M., DePaolis, R. A., & Keren-Portnoy, T. (2014). The role of production in infant word learning. *Language Learning*, *64*(s2), 121-140. https://doi.org/10.1111/lang.12058

Vihman, M. M., Kay, E., de Boysson-Bardies, B., Durand, C. & Sundberg, U. (1994). External sources of individual differences? A crosslinguistic analysis of the phonetics of mothers' speech to one-year-old children. *Developmental Psychology*, *30*(5), 651-662. https://doi.org/10.1037/0012-1649.30.5.651

Vihman, M. M., Nakai, S. & DePaolis, R. A. (2006). Getting the rhythm right: a cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, K. Best, D. Whalen & S. Anderson (Eds.) *Papers in Laboratory Phonology VIII: Varieties of Phonological Competence* (pp. 341-366). Cambridge: Cambridge University Press.

Vihman, M. M., Ota, M., Keren-Portnoy, T., Lou, S. & Choo, R. Q. (2022). Child phonological responses to variegation in adult words: A cross-linguistic study. *Journal of Child Language*, 1-28. https://doi.org/10.1017/S0305000922000393

Vihman, M. M., Ota, M., Keren-Portnoy, T., Choo, R. Q, & Lou, S. (2023). A challenge to whole-word phonology? A study of Japanese and Mandarin. *Language Learning and Development.* https://doi.org/10.1080/15475441.2022.2149401

Vitevitch, M. S. & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481-487. https://doi.org/10.3758/bf03195594

Vitevitch, M. S. & Luce, P. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, *52*(2), 193-204. https://doi.org/10.1016/j.jml.2004.10.003

Walley, A. C. (1993). The Role of Vocabulary Development in Children's Spoken Word Recognition and Segmentation Ability. *Developmental Review*, *13*(3), 286-350. http://doi.org/10.1006/drev.1993.1015

Wan, I.-P. & Jaeger, J. (1998). Speech Errors and the Representation of Tone in Mandarin Chinese. *Phonology*, *15*(3), 417-461. https://doi.org/10.1017/S0952675799003668

Wang, C. Y.(2011). *Children's Acquisition of Tone 3 Sandhi in Mandarin*. PhD thesis, Michigan State University, East Lansing, MI.

Wang, W. S.-Y. (1969). Project DOC: Its methodological basis. *Journal of the American Oriental Society*, *90*(1), 57-66. https://doi.org/10.2307/598431

Wang, W. S-Y. & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, *10*(3), 629-636. https://doi.org/10.1044/jshr.1003.629

Wang, W. S-Y., Li, K. P. & Brotzman, R. L. (1963). Inventory of Mandarin syllables. *Papers on Linguistic Analysis*, *6*. 27-36.

Wang, X. & Huang, J. (2016). Non-word repetition performance in Mandarin-speaking preschool children with and without Specific Language Impairment (SLI). *Chinese Journal of Applied Linguistics (Quarterly)*, *39*(3), 337-375. https://doi.org/10.1515/cjal-2016-0022

Wauquier S. & Yamaguchi N. (2013). Templates in French. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches, cross-linguistic evidence* (pp. 317-342). Cambridge, UK: Cambridge University Press.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*, *7*, 179-211.

Wee, L. (2013). Language policy in Singapore: Singlish, national development and globalization. In E. Erling & P. Seargeant (Eds.), *English and development* (pp. 204-219). Clevedon: Multilingual Matters

Werker, J. F., Hall, D. G., & Fais, L. (2004). Reconstruing U-shaped functions. *Journal of Cognition and Development*, *5*(1), 147-151. https://doi.org/10.1207/s15327647jcd0501_15

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wewalaarachchi, T. D., & Singh, L. (2016). Effects of suprasegmental phonological alternations on early word recognition: Evidence from tone sandhi. *Frontiers in Psychology*, *7*, 1-14. https://doi.org/10.3389/fpsyg.2016.00627

Wewalaarachchi, T. D. & Singh, L. (2020). Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin. *Journal of Experimental Child Psychology*, *189*, 1-23. https://doi.org/10.1016/j.jecp.2019.104698

Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English-Mandarin toddlers. *Journal of Experimental Child Psychology*, *159*, 16-33. https://doi.org/10.1016/j.jecp.2017.01.009

Wiener, S., & Turnbull, R. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*, *59*(Pt 1), 59-82. https://doi.org/10.1177/0023830915578000

Wong, James. (2003). The rise and decline of Cantopop: a study of Hong Kong pop music. PhD thesis, University of Hong Kong. [cited in Lim, 2007]

Wong, Jock. (2003). The reduplication of Chinese names in Singapore English. *RASK, 19*, 45-85.

Wong, J. (2004). The particles of Singapore English: a semantic and cultural interpretation. *Journal of Pragmatics*, *36*(4), 739-793. https://doi.org/10.1016/S0378-2166(03)00070-5

Wong, J. (2006). Contextualizing *aunty* in Singaporean English. *World Englishes*, *25*(3-4), 451-466. https://doi.org/10.1111/j.1467-971X.2006.00481.x

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics*, *40*(1), 141-151. https://doi.org/10.1016/j.wocn.2011.10.005

Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *Journal of the Acoustical Society of America*, *133*(1), 434-443. https://doi.org/10.1121/1.4768883

Wong, P., Fu, W. M. & Cheung, E. Y. L. (2017). Cantonese-Speaking Children Do Not Acquire Tone Perception before Tone Production: A Perceptual and Acoustic Study of Three-Year-Olds' Monosyllabic Tones. *Frontiers in Psychology*, *8*, 1450. https://doi.org/10.3389/fpsyg.2017.01450

Wong, P. & Strange, W. (2017). Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PLoS ONE*, *12*(8), e0182337. https://doi.org/10.1371/journal.pone.0182337

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. https://doi.org/10.1044/1092-4388(2005/074)

Woon, F. T., & Styles, S. J. (2021). *Red-Dot Baby-Talk Quiz.* Data repository available at https://doi.org/10.21979/N9/DYCWSY, DR-NTU (Data)

Xiandai Hanyu Cidian. (1979). *Modern Chinese Dictionary*. Beijing: Shangwu yinshu guan.

Xu, J. 徐杰 & Wang, H. 王惠. (2004). 现代华语概论 *Xiandai Huayu Gailun* (The Structure of Chinese Language: Characters, Words and Sentences). Singapore: Global Publishing

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61-83. https://doi.org/10.1006/jpho.1996.0034

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L. & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology*, *9*, 1-10. https://doi.org/10.3389/fpsyg.2018.00817

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: the role of tone information. *Language and Cognitive Processes*, *14*(5-6), 609-630. https://doi.org/10.1080/016909699386202

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. https://doi.org/10.1016/j.jml.2012.09.004

Yeung, H. H. & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234-243. https://doi.org/10.1016/j.cognition.2009.08.010

Yip, M. (1995). Tone in East Asian languages. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 476-494). Cambridge, Massachusetts & Oxford: Blackwell.

Yip, M. (2002). *Tone*. Cambridge University Press. https://doi.org/10.1017/CBO9781139164559

Yu, A. C. L. (2009). Tonal Mapping in Cantonese Vocative Reduplication. In I. Kwon, H. Pritchett & J. Spence (Eds.), *Proceedings of the 35th Berkeley Linguistics Society* (pp. 341-352). Berkeley, USA.

Yu, M., Mo, C., & Mo, L. (2014). The Role of Phoneme in Mandarin Chinese Production: Evidence from ERPs. *PLoS ONE*, *9*: e106486. https://doi.org/10.1371/journal.pone.0106486

Zamuner, T., Gerken, L. & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language*, *31*(3), 515-536. https://doi.org/10.1017/s0305000904006233

Zamuner, T., & Thiessen, A. (2018). A phonological, lexical, and phonetic analysis of the new words that young children imitate. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique*, *63*(4), 609-632. https://doi.org/10.1017/cnj.2018.10

Zhang, C. H. (1988). 广州话对新加坡华语的影响 *Guangzhouhua dui xinjiapo huayu de yingxiang* (The Influence of Cantonese on Singapore Mandarin). *Huawen yanjiu (Chinese Language Research)*, 2, 36-45.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(1), 153-201. https://doi.org/10.1017/S0952675710000060

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, *10*(6), 545-600. https://doi.org/10.1080/01690969508407114

Zhu, H. (2002). *Phonological development in specific contexts: studies of Chinese-speaking children*. New York: Multilingual Matters.

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, *27*(1), 3-42. https://doi.org/10.1017/s030500099900402x

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes*, *11*(6), 589-596. https://doi.org/10.1080/016909696386980