Early Phonological development in Mandarin: An analysis of prosodic structures, segments and tones from babbling through the single-word period

Shanshan Lou

PhD Thesis

Language and Linguistic Science

University of York

November 2020

ABSTRACT

This study presents an in-depth analysis of phonological development in Mandarin-learning infants, from babble through the early word period. Different from previous Mandarin studies which focused only on the order of acquisition and accuracy in production (e.g., Zhu, 2002), we investigate segment and tone acquisition as part of an emergent system. This thesis presents three perspectives on our central research question: Is an ambient language effect detectable on the use of tone in Mandarin- and English-learning infants' pre-linguistic vocalizations (babbling period)? What segmental advances and challenges will Mandarin-acquiring children have in their early word production? What tonal advances and challenges will Mandarin-acquiring children have in their early word production? The role of the input is also considered, with a prosodic and segmental analysis of infant-directed speech.

The contributions of the thesis are in three aspects. The results provide evidence that Mandarin-acquiring infants begin to shift to language-specific tone patterns already within the babbling period, by showing that a considerably larger number of vocalizations with Tone 1 are produced by Mandarin-learning infants than by English-learning infants. English- and Mandarin-learning children's early cross-linguistic commonalities can be ascribed to biological foundations, such as physiologically natural uses of the laryngeal muscles.

The analyses were made on two lexically defined points: the 4-word point (4wp) and the 25-word point (25wp), in the single-word production period. Mandarin children's prosodic structures are mainly of the form CV and CVCV at the 4wp and are broadly similar to those of children learning other languages. When the lexicon has grown larger, at the 25wp, they begin to attempt to produce more challenging structures, including words with diphthongs, codas and more than one consonant type, and their productions are distinguishable from those of children learning other languages. As for segmental learning, stops and nasals characterise Mandarin children's pattern at the 4wp; velars, fricatives and affricates occur less but those produced all tend to match the targets. The preference for stops and nasals is still found at the 25wp, but the use of fricatives and affricates increases and stands out from children learning other languages. Mandarin input speech also demonstrates a high frequency of fricatives and affricates.

Separate developmental paths were found for each tone in the single-word production period.

In monosyllables, Mandarin-learning children mainly use Tones 1 and 4 at the 4wp, but not always as matches to target; this preference is still seen at the 25wp, but with more accurate use. Tone 2 is produced highly accurately in a few items at the 4wp and more often but less accurately at the 25wp. Tone 3 emerges only at the 25wp, in few forms but always accurately. In general, the proportion of Tone 1 and Tone 4 use in monosyllables decreases as Tone 2 begins to be used. At the 4wp Tone 3 is replaced by either Tone 1 or Tone 4 but Tone 1 no longer substitutes for Tone 3 at the 25wp; instead, the perceptually similar Tone 2 is the main substitute. Possible reasons for the uniqueness of tone in language acquisition are discussed.

Table of Contents

List of Tables

List of Figures

ACKNOWLEGEMENTS

I am deeply indebted to my supervisors Professor Marilyn Vihman and Dr. Tamar Keren-Portnoy for their intellectual support and encouragement, which helped me overcome many challenges throughout the work. I am grateful to both of them for their guidance and valuable comments through this PhD.

I would like to thank all the Chinese parents and babies who took part in my field studies and showed me generous support in this research and the cooperation that made this thesis possible. I would like to express my profound gratitude and countless thanks for the education, friendship and encouragement I have received from many scholars and peers I have been privileged to know over the course of my academic training. Thanks must go to Catherine Laing and Amanda Cardoso for taking the time to show me the pitch analysis in PRAAT and go over some of my infant utterances. I owe special thanks to Wenling Cao and Cui Wang for their assistance on transcription reliability, and to Mao Li for the technical support on all adult experiments through this PhD.

I would like to express my heartfelt gratitude to my parents and my daughter for their unfailing love, understanding and support, without which this PhD would have never come to fruition. I am deeply grateful to my mother for being there and supporting me through the dark hours and the hardest times, even during her period of illness.

DECLARATION

I declare that the content of this thesis is my own work, and any content from external sources has been acknowledged through explicit referencing and quotations where appropriate. No part of this thesis has been submitted for examination at this or any other institute for another award.

1. INTRODUCTION

In the first 18 months of life, infants show impressive feats of learning in their native language, moving from having no linguistic system to the beginnings of a system. Mandarin Chinese, a tone language, is the most widely spoken language in the world (Yip, 2002), yet Mandarin-acquiring infants' early phonological development has so far received little attention. Over the past 20 years studies of phonological development in Mandarin have largely addressed aspects of infant speech perception rather than production. Researchers generally agree that infants show discrimination of the tone categories that are relevant for their language (Mattock and Burnham, 2006; Tsao, 2008), similarly to vowels (Polka & Werker, 1994) and consonants (Werker & Tees, 1984), even as early as four to six months (Yeung et al., 2013). Although tone categories seem to emerge early in perception, this leaves open the question of when infants attune to the native tone inventory in production, as appears to be the case for nascent vowel and consonant categories.

The notion of babbling drift is still debated – that is, whether ambient language influence is already sufficiently strong to shape infants' vocalizations in the babbling period. It is believed that infants gain a more solid native-language phoneme inventory from the period of word production. It is a fascinating question whether there is babbling drift in terms of tone in Mandarin-acquiring infants. This thesis will provide evidence that Mandarin-acquiring infants begin to shift to language-specific tone patterns already within the babbling period. We will argue that children's early cross-linguistic commonalities can be ascribed to biological foundations, such as infants' immature production capacities, rather than to linguistic foreknowledge.

Over the past two decades there has been a surge of interest in how language is acquired as a system. The most recent production study on Mandarin drew on generative phonology to explain development in terms of the order of acquisition of features or rules which determine child performance (Zhu, 2002). The aim of this study is to investigate Mandarin-learning children's phonological development holistically, in terms of prosodic structures and both segmental and tonal categories. Our approach draws on exemplar theory and usage-based phonology, which claim that children's phonology emerges principally from what they hear and attempt to say; this approach will underlie our attempt to understand Mandarin-learning infants' phonological development from babbling to word use. This thesis uses naturalistic

home recordings to investigate: first, the origin of lexical tone use in the babbling period; second, the strengths and challenges of phonological acquisition in Mandarin-learning children, in comparison to what has been shown in previous studies of children from other linguistic backgrounds; third, factors that influence phonological development. The cross-linguistic analysis demonstrates three elements that influence phonological acquisition: (i) the extent of universality in development, due to biological constraints, (ii) language-specific properties or salient patterns in the ambient language environment and (iii) individual preferences.

The acquisition of lexical tone offers insights into the development of a linguistic system that are different from what can be gained through studying the development of lexical stress or intonation, which are properties of prosodic systems that do not make lexical use of tone. Cross-linguistic comparisons between Mandarin and other languages will advance our understanding of the mechanisms behind child phonology phenomena and developmental processes.

This thesis is based on both cross-sectional and longitudinal investigation. We will begin by reviewing theoretical models and the trajectory of phonological development in the first two years; we also provide an introduction to Mandarin phonology. Finally, we present the specific research questions and contents of each chapter in this thesis.

1.1Theoretical approaches to child phonology

Phonological development is based not only on linguistic aspects of acquisition but also on the psychology behind the learning process, that is, the mechanisms of system creation in the child. Chomsky (1964) claimed that language is a human genetic inheritance, known to the infant in advance of experience. He denies any active role for the child in language development but asserts that the infant's prior knowledge of linguistic structure accounts for the speed of acquisition. Since Chomsky and Halle (1968), generative phonology has emphasized the abstract nature of phonology, such as distinctive features. Supporters of this approach believe that phonetic knowledge must be derived from inherently available general principles or general grammar (see Burton-Roberts, Carr & Docherty, 2000).

Since the 1980s a series of changes have taken place in the fields of adult and child phonology. Scholars have begun to shift their attention from rule-based and segmental

phonology to output forms and usage (Vihman, 2014). In the fields of phonology, phonetics and morphology, the analysis of stripped-down, atomic units has been transformed into a more comprehensive study of words, schemes and patterns (Pierrehumbert, 2016). Optimality Theory is widely used as the preferred framework for phonological research (Scheer, 2012). Prosodic analysis (Menn, 1983) and analysis that relies on structure, schemas or whole-word units (Croft, 2001) are used in child phonology to understand acquisition in an emergentist way. In addition, the concept of exemplars in psychology, which rejects the normalization of auditory forms in the process of speech perception, began to gain attention in the study of speech development in the 1990s (Jusczyk, 1992; Johnson, 1997). A number of studies established the role of statistical or distributional learning, which is considered a basic mechanism for acquiring structural knowledge (e.g., Saffran, Aslin & Newport, 1996). These studies have increased our understanding of the structure of infant language learning. The acquisition of phonology or phonological development is rooted in the speech-motor and perceptual systems. Many issues remain unresolved, but a combination of implicit statistical learning and item learning seem sufficient to account for the linguistic knowledge that children attain.

Biological model

The biological approach was first applied to phonological acquisition by Locke (1983). Rejecting the idea that knowledge of linguistic structure is available from the beginning, he sees children's acquisition of phonology as influenced by articulatory and perceptual constraints. A small core of consonants was found in the babbling of children learning many different languages; this universal babbling repertoire is related to physiological, perceptual and cognitive components. These anatomical and neuromotor factors determine the match between the infant's own patterns and patterns of the adult language. Subsequently, Kent (1984) claimed that language is constrained by several factors, such as auditory perception, speech production and cognitive experience. Genetic factors are represented by both universal perceptual categories and ambient language constraints on early production. Speech motor control is an important factor affecting infants' performance and is acquired in a continuous but nonlinear process. MacNeilage and Davis (1990) proposed that the acquisition of motor representation and control derives from a universal motor base: the rhythmic alternation between open and closed jaw (mandibular oscillation). This is an infants' first adult-like syllable production. The 'content' of syllables is the mechanical consequence of lip and tongue placement with respect to the moving mandible and constitutes a limited number

of fixed consonant-vowel sequences. Certain consonant-vowel associations appear to obtain first; children are gradually released from the constraints of their early sound patterns. A series of analyses of these perceptuomotor constraints on early phonological development have been conducted on children learning the same or different languages (e.g. Vihman, 1992).

Exemplar model

The exemplar model is one of the emergentist models in phonological development. The model claims that infants start to acquire a phonological system by retaining many individual instances from about 8 months of age; the instances they take in are words or word-like sequences (Menn et al., 2013). Jusczyk (1992), who outlines the differences between prototypes and exemplars, first brought the idea of exemplar representations into the child phonology literature. In contrast to other cognitive models, which argue that only abstract representations of categories are stored in memory, Jusczyk believes that individual episodes are also stored and are used to represent the entire category in retrieval. That is, a word can be extracted by referring to previous events. In word production, this "instance-based model could provide automaticity in skill learning" (Logan, 1988, p.46) This means that infants who are often exposed to a specific vocalization, especially one that they themselves produce, can gain automatic access to a routine; infants can activate the routine on hearing roughly similar phonological sequences in input speech.

However, not all utterances can be easily recorded as episodic traces; infants face the challenge of retention in the early word-learning period. Experience cannot modify previously stored traces but can only add a new trace to memory. Each trace is activated based on its similarity to the input. Assuming a lack of pre-existing linguistic knowledge, infants' early lexical representations would be very incomplete and variable. Only the overall stress pattern and a few phonetic features would be stored (Jusczyk, 1992). Following this idea, a series of experiments were conducted and showed that children neglect the changes in unaccented syllables (Hallé & Boysson-Bardies, 1996; Vihman et al., 2004).

Exemplar representations are prioritized as rich memory representations because they contain all the potential pieces of information which are important in a linguistic experience. Therefore, when a child learns words, each new experience will influence the learning process; due to the accumulation of new and innovative traces the whole memory system

functions in the most effective way, although this process will not so much affect adults and remains an open issue (Bybee, 2010).

Based on the exemplar theory of cognition, Pierrehumbert (2003) proposed two types of organising processes in the phonological system: bottom-up processes, in which the exemplars experienced are grouped together based on similarities between them and categories emerge from dense groupings of similar items, and top-down processes, in which novel exemplars are categorised on the basis of previously established categories. Vihman et al. (2009) claim that the emergence of templates as categories is the first top-down effect in a child's own production. That is, dense categories based on the child's own forms work as 'magnets' for new exemplars, which become modified in memory and production.

Linked-attractor model

In usage-based models, practice affects grammar, and phonology arises from accumulating experience. The Linked-attractor model (Menn, Schmidt & Nicholas 2009) is a usage-based approach which asserts that children are not developing rules related to underlying forms but are mapping the input forms to their own patterns. The model generally consists of the three-components of input, storage and output. The input is what the child hears and registers, storage is the place where the information is kept, and production is what the children say. There are three kinds of attractors in a child's phonological knowledge: perceptual/input, production/output, and inter-modal attractors, which are the mappings between input and output. These attractors are built up gradually by experience with hearing or seeing (e.g., labials), understanding, and speaking; therefore, links between representations get stronger each time they are used.

Infants show word-form recognition by the end of the first year (Jusczyk 1997). The Linked-attractor model claims that infants' first perceptual or auditory/acoustic attractors are the result of both the speech of adults and their own babble vocalizations. Each target attractor is the representation of a separate adult lexical item. This process involves both the situational meaning and the linguistic meaning and the importance of social responses in the affective association. The link starts from the daily routine, in which words are used in recurring situations. The Linked-attractor model is exemplar-based, so it emphasizes the role of frequency in both input targets and output forms. From a production perspective, repeated jaw or tongue movements and vocal fold vibration create the infant's early vocalizations

(Vihman 1996); at a later period, the accumulation of word production experiences, including imitation, helps to store the output lexical items. Existing words in the early lexicon also attract new words. Each time a child attempts to say a word, the links are strengthened between the articulatory representation of how the child said it and the actual sounds that resulted from the child saying it.

The Linked-attractor model (Menn et al., 2009, 2013) considers that the representation of a word's form develops continuously over time in strength, precision and accessibility. That is, the use of language can affect word formation. Strength and precision are achieved through repeated perception, production or the mappings between them. The multimodal attractors reflect various aspects of production and perception. This model reconceptualizes 'representation', which has long been psycholinguistically necessary for fully understanding language development. And these models allow us to bring frequency data to bear on development without ignoring the equally important contributions of linguistic structure.

The usage-based model of language development states that children build their language with the help of the concrete construction of individual words or frames and they develop language or speech based on the sounds they hear and use. Menn and her colleagues found a strong connection between babble and early words. It is found that many children produce a word that 'falls into' the late babble articulatory attractors. This implies that late babble attractors are essentially articulatory routines for early words. The role of the child's 'rule' is essential as it links 'attractors' and creates a link between the adult form and the child's production. Thus, it can be said that babble practice and infant experience of hearing the same sounds and speech forms in their daily routines, together with early word use, lead to the construction and development of input and output lexicons. A dense network of connections is gradually developed between the word forms of the child and that of the adult targets, between the phonological forms of the word and similar-sounding words, and between the meanings in situational contexts related to those word forms. This model is the foundation for children acquiring the adult language system.

Whole word phonology

Whole-word phonology was brought to the attention of researchers from the 1970s (Waterson, 1971; Ferguson & Farewell, 1975; Priestly, 1977; Macken, 1979). The approach is related to other usage-based models but emphasises individual differences and nonlinear

phonology. The idea of whole-word or holistic representation was first identified by Waterson (1971). In detailed analyses of her son's early word patterns, Waterson found that the child's words have a holistic relationship with the target, rather than segment-by-segment, and appear to constitute an individual systematization of the child's production. The child reproduces only the most salient acoustic features shared by the adult words in his repertoire. For instance, the child not only uses palatal [ɲ] to substitute [n] but also applies a disyllabic nasal structure in response to adult words that contain a nasal in the stressed syllable (e.g. [ɲaɲa] for *another*; [ɲẽːɲẽ] for finger; [ɲeːɲeː] for *window*).

Ferguson and Farwell (1975) supported this idea by showing that children's early word forms are simple; the words they use are variable rather than based on consistent rules, but each one is very close to the target. The simplicity and accuracy of the first words indicates that there might be 'phonological selectivity' in early word use. In other words, the children first try to use the words or word forms that they have already mastered in a certain way. That is, they choose from their babble repertoire; children's early words are found to be closely related to the vocal forms they produced in babbling (Vihman et al, 1985). Stoel-Gammon (2011) also found that lexicon and phonology are beneficial to each other; vocalisation and the phonetic skills developed in babbling are the basis of word learning. The selectivities or unconscious choices of children's first words depend on the frequency and prosodic saliency of certain words in the input language as well as on children's own articulatory and memory capacities – that is, their ability to remember and retain word forms in related situations or context.

Macken (1979) defended this idea by emphasizing that the word is the primary organizational unit in a child's early phonology. Macken observed one child's phonological development over a 10-month period. The child matched the targets in a holistic way and consonantal contrasts were learned through evolving word patterns.

Prosodic structure is the starting point of holistic analysis. The term 'prosodic structure' has been used in many different ways in the phonological literature: the accent or stress pattern of words, for example, strong-weak or weak-strong stress on disyllables (e.g., Jusczyk, Cutler & Redanz, 1993); the overall rhythm pattern of the language (e.g. Mehler, Dupoux, Nazzi & Dehaene-Lambertz, 1996); the units of the prosodic hierarchy, for example, the syllable, foot, prosodic word, phonological phrase, intonational phrase and utterance (Selkirk, 1980a); only the lowest level of subparts of syllables like onsets and rimes or moras (Fikkert, 1994) or

early word forms attempted by children and the word production mode commonly used by children in terms of the overall word shape (Vihman, 2019). The idea is that prosodic structure is the starting point for infants to learn the grammatical structure of their language (Gleitman and Wanner, 1982), because they are sensitive to the rhythms of their language very early and can extract acoustic features from running speech (see Jusczyk, 1997, for experimental evidence). Children's initial representations of prosodic structure are related to sensorimotor and cognitive factors, including the perception of salient elements in the input speech.

Vihman and her colleagues interpret the whole word approach to language production as a template-based approach. The prosodic analyses of McCarthy (1979) inspired the first use of the term 'template' in child phonology, by Menn (1983). The researchers mentioned earlier tended to analyse the data of a single child. To establish the validity of template structures, Vihman and Croft (2007) examined diary studies of children learning several languages. They viewed the phonological form as the simplest symbolic unit and also a possible 'usage event'. The templates each child develops are the first step in phonological organization. After producing their first words, children develop not only a range of phonological categories from the known word shapes but also a small number of phonological templates. The VCV templatic pattern is observed in some children each learning Estonian, Finnish, French or Welsh, for example. In two early stages of phonological development (the transition from babble to words and the emergence of templates), a shift results from the interaction between children's input and output forms. Templates are children's responses to the challenge of producing a distinct consonant or vowel in different positions or syllables. Templates are emergent from dynamic self-organization.

Different kinds of evidence provide different principles for identifying templates in early language production. In Vihman's studies (Vihman, 2014; Vihman and Wauquier, 2018) a template is mainly regarded as a prosodic structure, referring to children's early creation of entire word shapes. It comprises phonological highlights and prosodic structures that are typical of the ambient language, specifically, syllabic patterns like CV monosyllables, CVCV disyllables and VCV or CVC, or segmentally specified patterns like consonant harmony. The words form preferred structures or patterns that are modified by the child according to their capacity. There are two types: specific phonological patterns which fit many of the words that the child attempts - termed 'selected' - and extended patterns that are less close to the child's

template - termed ‘adapted’. Vihman (2019) investigated children learning each of six languages, showing similarities among all these languages and also that each language has its own characteristic: The most used pattern, disyllabic consonant harmony, is evidenced by at least one child in each language; the less commonly used VCV pattern is found in four children learning three languages, Finnish, French and Italian; reduplication is only found in one Finnish child. Individual differences within language groups are also in evidence, for example, a templatic use of a palatal pattern in one child and nasal codas in another child. On the other hand, in Kehoe (2015), infants implement strategies including preferences for certain sounds or sound classes and extensive use of reduplication. Kehoe (2015) found the salient use of initial consonants or specific consonant codas as templates in both German- and Spanish-learning children: For example, /h/-initial patterns are frequently used in both selected and adapted forms in German children and /k/-initial patterns are seen in Spanish children; a monosyllabic final-labial pattern is identified in two children, a disyllabic final-/n/ pattern in one child and a final-/t, d/ pattern in one child.

The essential puzzle of child language production is the origins of the linguistic system. However, the advances afforded by exemplar theory, in which each perceived utterance takes its place in experienced speech, providing a model of representation, and usage-based models, in which vocal production affects processing in both production and perception, give children a rich platform to build on.

1.2 The trajectory of phonological development: From canonical babbling to word production in the first two years

A child’s passage into human culture starts with advances in language. Caregivers regularly recognize this achievement after hearing their child’s first words. In the context of the early programmatic ideas of the Prague school Jakobson (1940) proposed that the sounds produced in babble are phonetically meaningless and unconnected to the language being learned, and general ‘laws of irreversible solidarity’ are responsible for governing the languages of the world synchronically. According to his account, children’s first words are completely different from babble; children’s acquisition of ‘phonemes’ follows a strict order of emergent oppositions. Therefore, the use of phonemes, which are simple, straightforward and stable, is more reliable as they are engraved in the memory and can be realized quickly.

On the other hand, Brown (1958) stated that babbling is a period when a child’s vocalizations

increasingly show effects of the ambient language, which he termed “babbling drift”. Brown proposed that children’s pre-linguistic vocalizations are affected by their experience of the ambient language before they begin associating the word forms and meanings in concrete situations. This view has been supported by studies using acoustic analysis of the vowel space (Boysson-Bardies, Hallé, Sagart & Durand, 1989: Arabic, British English, Cantonese, French), place and manner of consonants (Boysson- Bardies & Vihman, 1991: American English, French, Japanese, Swedish), and lexical tones (Lou, Vihman, & Keren-Portnoy, 2018: British English, Mandarin). The question of babbling drift has also been contested in some studies using adult listener judgments as to linguistic influence on babble. Listeners cannot always successfully distinguish whether the vocalizations infants produced were from the listener’s own language or other languages (e.g., Engstrand, Williams & Lacerda, 2003; Lee, Jhang, Chen, Relyea & Oller, 2017).

Investigation of pre-linguistic vocalizations was conducted over the last quarter of the twentieth century. Contrary to Jakobson’s view that babbling is not essential and cannot serve as production practice, the importance of babble production on children’s language development is now well-established. First, babble provides a direct role for supporting early social interactions in phonological development and word learning. In many cultures babbling arouses the attention of caregivers, and caregivers’ responses can provide useful models for word learning. Goldstein, King, and West (2003) found a parallel to early vocal development in the study of birds. Only after young birds realize the value of sound as a tool does early sound become a channel of communication: The similarities between bird and human vocal development lie in the passing of some kind of primitive social information. Infants’ vowel quality and consonant-vowel transitions become more speech-like when contingent caregiver feedback is given. Infants also learn word-object associations more easily when adults respond contingently to their object-directed vocalizations with object labels (Goldstein, Schwade, Briesch & Syal, 2010). Second, babble is necessary and is a significant predictor of later word production. Some premature infants cannot babble or vocalize immediately after birth because their vocalization is restricted by the presence of a cannula inserted in the vocal tract to support the breathing process. These babies are prevented from vocalizing by the cannula, but no matter when the cannula is removed, they always babble for one to several weeks before producing their first words (Bohm et al., 2010). As McGillion et al. (2017) review from various studies, in the case of poverty, deafness, Down syndrome or prematurity children may show some delay for typical

babbling. Those children who do not babble in the typical time, that is, by 24-30 months, almost all speak later than typically developing children, since they were unable to profit from babbling practice. Therefore, stable and consistent consonant use in babble is considered to be a precursor of first word production.

Babbling is made possible by dramatic postnatal changes in vocal tract shape and by the help of the maturation of neurophysiological control. Like hand banging and kicking, babbling is type of maturing motoric skill with a rhythmic base (Thelen, 1991). Canonical syllables are observed after a period of use of reflexive or 'vegetative sounds', cooing and vocal play. The first use of supraglottal consonants can typically be identified from age six to eight months in canonical babbling within a rhythmically repeated syllabic frame like [bababa] or [dadada] (Oller, 1980). Canonical babbling is considered a milestone in vocal production in the first year of a child's life because it resembles adult syllables. Davis and MacNeilage (1990) argue that mandibular oscillation and cyclic jaw movement are the key movements for children to produce adult-like speech; and during this period, the syllables show a strong trend of C-V association. That is, a coronal consonant is regularly followed by a fronted vowel; the less frequently produced velar consonants tend to be followed by a back vowel; labials include no tongue movement and are commonly followed by central low vowels. In other words, Davis and MacNeilage (1990) found that when canonical babbling emerges, children can control the forward movement of the tongue, which produces the auditory effect of consonants.

First words and the articulatory filter

As more recognizable consonants gradually appear, children make the transition from babbling to words (Stoel-Gammon, 1992). The intentional use of gestures to communicate meanings and the capacity to produce stable vocal forms are precursors for language use (Vihman, 2014, ch. 2). A child's first word typically occurs around the time of the first birthday. In this first-word period, the adult-based word forms children produce are restricted to the social and action context of children's routine life. Also, a child's own production plays a role on the perception of input by serving as an articulatory filter (Vihman, 1993). The 'articulatory filter' is a mechanism by which children experience adult speech, selectively enhancing motoric recall of phonetically accessible words (Vihman, 1996). This means that the word forms that children best retain from the input are those that are close enough to their own vocal forms to provide a match. According to exemplar theory, similar forms create an 'echo' with known forms, facilitating the processing and retention of related

patterns heard in the input. In line with the neuroscience literature, the 'articulatory filter' idea claims that a child's own vocal production and perception are automatically coordinated in real time; production strengthens the network of neural connections.

The perception-action loop in the transition from babble to first words was examined by DePaolis et al. (2011), who showed that the sounds produced by children affect the sounds they attend to. These sounds, which have been termed 'vocal motor schemes' (hereafter VMS), reflect each child's own well-established phonetic skills (McCune & Vihman, 1987, 2001). McCune & Vihman operationalized VMS as the consistent use of a supraglottal consonant in babbling or words in at least three out of four consecutive monthly recordings. Children were tested to listen to speech forms both with their VMS and without and were found to show a preference for patterns with their own VMS rather than patterns without (DePaolis et al., 2011). Westermann and Miranda (2004) explained how a connection between production and perception could develop in the mind. That is, repeated coactivation between items in the auditory and the motor maps in the brain leads to the development of a strong sensorimotor coupling. Therefore, perceived sounds activate the motor response, so that they can be reproduced.

Meanwhile, familiar words play a similar role in arousing children's memory and attention. Given that a large number of words are used repeatedly in the daily life of infants, many possible words will be suggested for expression. The construction and familiarity of these words results from the child's intrinsic interest and their frequency of occurrence (Vihman, 2018). These high-frequency words include greetings, animal names, and items in the daily diet. The frequency of occurrence of words and their rhythm help infants to recall them. Infants have been found to respond differently by 11 months to high-frequency word forms which are familiar from their everyday experience as compared with unfamiliar word forms (e.g., Hallé & Boysson-Bardies, 1994; Vihman et al., 2004).

Therefore, the first words are formed by the match between a prelinguistic 'articulatory routine' and salient and frequent adult words. Children just map input forms to their vocal schemes in this early word learning; meaning is not necessarily implicated.

Templates and early systematicity

Following a period of consistent phonetic practice along with the maturation of articulation and memory, infants start to produce more words. With the growth of infants' lexical experience, stronger semantic links are created. Existing words facilitate the learning of new words. Children now begin to produce words with generalized meanings in cross-situational uses rather than individual item-based words; this has been termed 'referential word use' (Vihman, 2017, 2018). This reveals the first step in establishing a phonological system. Vihman and Velleman (2000) consider that phonology is emergent from the interaction between articulation and perception and from children's own generalisations across both the input and the output forms they experience. Generalisations are each child's individual solution to the mismatch between the input and their own production; such generalisations lead to the patterns which have been termed word templates. It has been observed that after the first word period of relatively accurate production, many children begin producing word forms with a particular structure. Children pick up on word patterns that fit their existing production routines. Templates are early cognitive categories which account for mismapping after the period of early accurate production. Children extend the phonological patterns in the words they are already producing to new, more challenging word targets and similarly extend the meaning of words beyond particular referent objects and events. The words children use show a regression in accuracy but reveal the existence of inter-word connections. That is, templatic patterns reveal similarities across the child's word shapes.

1.3 The role of the input language

The mechanism of language learning remains in dispute. One account holds that a human baby can only learn a language due to an innate capacity for language, termed Universal Grammar (UG). Supporters of UG conclude that no learning is possible without a priori structure and hold that learning is minimal. However, it is also true that although there are some universal biological constraints, phonology, the lexicon and the grammar vary from language to language. Infants can use statistical data to learn types of language structures. The usage-based model of naturalistic language acquisition holds that infants learn to use language on the basis of linguistic experience. Therefore, frequency plays a role in the learning process. There are two relevant types of frequency: input frequency, or the frequency of occurrence of forms children hear in the ambient language, and output frequency, or the experience children gain by articulating particular words (this would enhance their accuracy in producing these words: Keren-Portnoy et al., 2010).

Infants have been found to be sensitive to the statistical structure of the input as early as 6–9 months, and acquisition follows a scheme on the basis of the input. The emergent knowledge can be seen to be rooted in the exemplars of similar forms that children have frequently heard, including information for both form and meaning (Pierrehumbert, 2003). The acquisition order or accuracy of different patterns depends on their relative frequency in the input. Therefore, the different characteristics of acquisition in different languages are due (at least in part) to quantitative differences in the input. The emergent approach demonstrates how linguistic categories can be induced based on distributional and frequency information in the input.

A pioneer study (Saffran, Aslin & Newport, 1996) tested infants' learning capacity on a statistical basis. Eight-month-old infants were exposed to a stream of three-syllable combinations from an artificial language, in random order, for two minutes. During the test trials, infants attended longer to rarely occurring three-syllable combinations (ones which did not constitute pre-established 'word units' in the training) than to familiar ones (the trained 'word units'). The result shows that infants are powerful learners who can pick up a structure from listening at as young an age as 8 months. Using natural speech syllables, Johnson and Jusczyk (2001) replicated Saffran et al.'s study using real-word syllables and provided further evidence of infants' ability to use statistical cues to segmentation.

Studies then moved towards establishing how statistical properties might be relevant for category learning. Maye et al. (2008) found that distributional information in input speech could facilitate 8-month-old infants' discrimination of a phonetic contrast. Children could discriminate the contrast if they were familiarized with a bimodal distribution, whereas they could not if they were familiarized with a unimodal distribution (see Figure 1.1). Statistical learning has been supported not only as applied in low-level processes like the categorization of speech sounds but also in higher-level processes like word or grammar learning. A word learning study showed that 12-month-old infants use distributional information to categorize novel words after brief familiarization (Mintz, 2006). Studies using the switch task asked 17-month-old infants to label objects and found that children did correct mapping more easily for statistical words (in which the test trials are the same as familiarized trials) than for words are not familiarized previously (Graf Estes et al., 2007; Werker et al., 1998).

Figure 1.1. The frequency distribution of input speech in bimodal vs. unimodal conditions (from Maye et al., 2008)

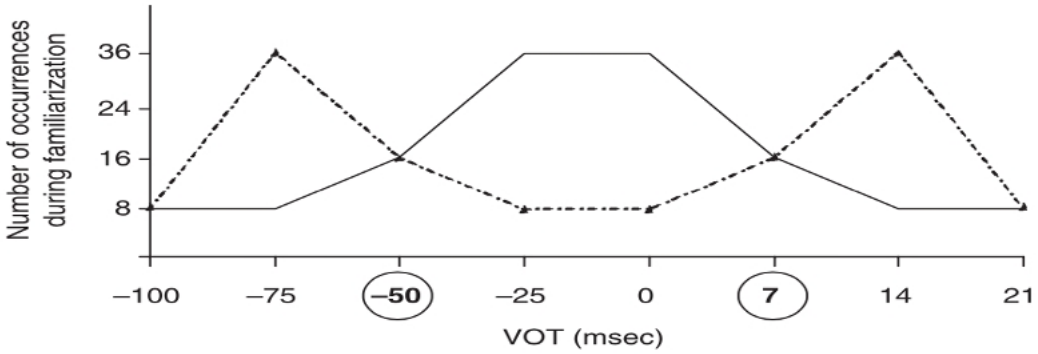


Figure 1.1 *Presentation frequency for continuum stimuli during familiarization phase. The frequency distribution for the Bimodal (Experiment 1) and Generalization (Experiment 2) conditions is shown by the broken line; the frequency distribution for the Unimodal condition (Experiment 1) is shown by the solid line. VOT values of –50 and 7 msec were used during the test phase.*

Frequency is also reflected in child language use (Bybee, 2010). Studies testing nonsense-word repetitions in English-learning children found that children produced high frequency sequences of phonemes more quickly and more accurately than low-frequency sequences (Munson 2001; Edwards, Beckman and Munson 2004). A corpus study investigating a German child's production rates also indicate that the distributional properties of adult language play an important role in shaping the child's language use. The child acquired the distributional properties gradually in the input structural domains (Behrens, 2006).

Input has been shown to play a role in the transition from babble to words. As well as having a general ambient language influence, contingent caregiver feedback to infant vocalizations also affects the phonological properties of babble. That is, the interactive adult responsiveness and the use of iterative social information refine infants' repertoires. In two studies, Goldstein, King and West (2003) and Goldstein and Schwade (2008) found that vowel quality and consonant-vowel transitions become more speech-like following contingent caregiver feedback. By manipulating mothers' responses to their infants' vocalizations, they showed that only contingent responses led to more speech- like productions in infants' babble. These findings reject the wholly 'egocentric' view of early vocalizations (Jakobson,1940/1968). Input speech also influences the production of a child's first words as a part of 'feedback loop'. That is, infants' word production may be shaped by

both the perception of the input speech and their own experience of practice in babbling.

1.4 Phonology of Mandarin

Putonghua or Mandarin (hereafter referred to as Mandarin) is the dominant language in China, although there are eight major dialects. Mandarin was prescribed in 1956 for nationwide use in schools and is widely used all over China nowadays in every activity, from broadcast news to commercial trades (Hou, 2002). Relative to English, which has more than 10,000 syllables, Mandarin Chinese has simple word and syllable structures: There are only around 400 distinct syllables (without counting tone) or 1,200 (when tones are considered: Chen et al., 2004; Xiandai Hanyu Cidian, 1979). However, Mandarin has a fairly large vowel inventory, along with phonemic use of tone and an average-sized consonant inventory (See Duanmu, 2000).

Consonants

Mandarin has a total of 22 consonants, with 21 (all except /ŋ/) allowed in initial position. Mandarin makes a distinction between aspirated and unaspirated consonants (Norman, 1988). There are six pairs of consonants with this distinction, all of them voiceless. Mandarin has a rich inventory of fricatives and affricates at different places of articulation. Table 1.1 provides the phonological consonant inventory.

Table 1.1. The consonants of Mandarin in IPA

	Bilabial		Labiodental	Alveolar		Retroflex		Palatal		Velar	
Stop	p	p^h		t	t^h					k	k^h
Nasal		m		n						ŋ	
Affricate				ts	ts^h	ʈʂ	$ʈʂ^h$	tɕ	$tɕ^h$		
Fricative			f	s		ʂ		ɕ		x	
Approximant				ɹ							
Liquid				l							

Vowels

Mandarin has a rich vowel system, with many diphthongs and triphthongs. Vowels (monophthongs, diphthongs and triphthongs) serve as the tone-bearing units in Mandarin. Mandarin has seven vowel phonemes: see Table 1.2.

Table 1.2. The vowels of Mandarin in IPA

	Front	Central	back
close	i y		u
mid		ɚ	ɤ o
open		a	

Note that /ɚ/ is a retroflexed central vowel. It occurs in isolation in some words (e.g., /ɚ2/ 'son'; /ɚ4/ 'two') or in rhotacisation of a suffix (replacing a nasal consonant, e.g. /wan2/ as [waɚ2] "play", Duanmu, 2000). The two high vowels, [i] and [u], usually occur as either the first or second vowel of a diphthong, combining with the mid and open vowels, i.e. [ai], [au], [ia] and [ua]. There are nine diphthongs /ai, ei, ou, au, ia, iɛ, ua, uo, yɛ/ and four triphthongs are /iau iou uai uei/. Table 1.3 provides examples.

Table 1.3. Examples of Mandarin words with diphthongs and triphthongs

	Vowels	Examples	Gloss
Diphthongs	ai	kai4	lid
	ei	pei1	cup
	ou	kou3	dog
	au	pau4	hug
	ia	tɕia1	home
	iɛ	ɕiɛ2	shoe
	ua	xua1	flower
	uo	ʈʂuo1	table

	yɛ	ɕyɛ3	snow
Triphongs	iau	tiau4	drop
	iou	tɕ^{h}iou1	autumn
	uai	k^{h}uai4	quick
	uei	tsui3	mouth

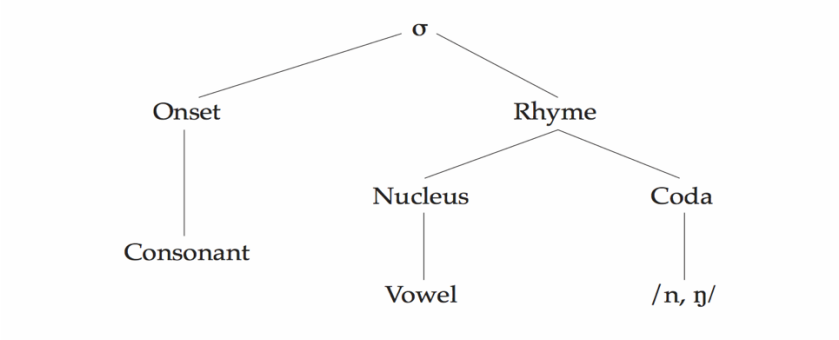
Syllable structure

Mandarin syllable structure is relatively simple compared with English. Syllable and word shapes include open syllables such as V, VV, CV, CVV, CVVV (VV = diphthong; VVV = triphthong), and closed syllables such as VC, CVC and CVVC. Sequences of consonants with syllable-final nasals in the first syllable can occur word-medially (e.g. [koŋ3loŋ2] 'dinosaur'. Some significant differences in syllable structure exist between Mandarin and European languages. The constraints can be classified and described as follows (Walton, 1976):

1. There are no consonant clusters. Only single consonant can occur in the initial and/or final position of a syllable.
2. Only nasals /n/ and /ŋ/ can appear in syllable-final position.
3. Generally no more than four speech sounds occur in a Mandarin syllable, and no more than three phones in the 'final' or nucleus, therefore only a single vowel or diphthong can co-occur with nasal coda.
4. The back nasal /ŋ/ is never employed as 'initial'.

The syllable structures of Mandarin are illustrated in Figure 1.2. Open syllables are more frequent than closed syllables in Mandarin. Of around 400 syllables, disregarding tone, C(G)V is the most common type in a large corpus (Da, 2007), at around 55% of the total, C(G)VC 42%, V 2%, and VC 1%. The vowels have the advantage of the consonants in oral Chinese; the frequency of occurrence of vowels is 54.2 (compare English 38%, Russian 43.3%: Duanmu, 2000).

Figure 1.2. The syllable structure of standard Mandarin Chinese (Zhu, 2002)



Phonotactic rules

The traditional framework of syllable structure is still widely used in the description of Mandarin Chinese phonology (Li & Thompson, 1981; Norman, 1988), which consists of three parts: initials, finals and tones. Based on traditional phonology the finals of Mandarin can be classified into four categories:

1. Finals with no medial vowel, called Kaikou: /a, o, e, er, ai, ei, au, ou, an, en, aŋ, eŋ/;
2. Finals with medial [i], called Qichi: /i, ia, ie, iau, iou, ian, in, iaŋ, iŋ/;
3. Finals with medial [u], called Hekou: /u, ua, uo, uai, uei, uan, uen, uaŋ, oŋ/;
4. Finals with medial [y], called Cuokou: /y, ye, yan, yn, ioŋ/.

There are 22 initials, including the 'zero initial', and 37 finals in Mandarin, which can make 814 possible combinations, but only about 410 syllables without lexical tones actually occur. Some strict phonotactic rules control the combination of an initial with a final to form a syllable. The distribution of the joint probability is not even and there are 12 impossible combinations. For examples, the dorso-palatal /tɕ, tɕʰ, ɕ/ can only be combined with the finals of Qichi and Cukou and never with the Kaikou and Hekou. In contrast, the apico-palatal /tʂ, tʂʰ, ʂ, ɹ/ have the opposite constraint. Details of the distribution are given in Table 1.4. The phonotactic rules that must be learned during the language acquisition process are treated as the internal information of the syllable structures (Zhang et al., 1982).

Table 1.4. The occurrence frequency of the initials and the joint probability of the initials in combination with the finals, %.

Initials	Finals			
	Kaikou	Qichi	Hekou	Cuokou
f 2.45	84.62	0	15.38	0
ʈʂ, ʈʂʰ, ʂ, ɻ 7.18, 2.75, 7.66, 1.94	75.13	0	24.87	0
t, tʰ 12, 3.53	59.04	20.87	20.09	0
k, kʰ, h 5.50, 1.83, 4.42	58.81	0	41.19	0
ts, tsʰ, s 3.01, 1.15, 1.08	54.81	0	45.19	0
p, pʰ, m 5.15, 0.98, 3.74	47.98	33.33	18.68	0
n, l 2.53, 5.69	46.38	41.58	10.17	2.03
ɸ 12.45	5.91	55.18	26.14	13.95
tɕ, tɕʰ, ɕ 6.98, 3.11, 4.86	0	78.73	0	21.27

Tones

In Mandarin, lexical tone is a key phonological category, in addition to vowels and consonants. Lexical tone is defined by f0 variations that contribute to meaning differences between words. Mandarin has four tones, which are generally transcribed using the 'Chao tone letter system' (Chao 1930/1980): Tone 1, which is high-level, has a mean pitch value of

55. The rising tone, Tone 2, starts at pitch level 3 and ends at level 5. Tone 3 is a falling-rising tone with pitch value 214. Tone 4, with a falling contour, starts at level 5 and ends at level 1 (Table 1.6). To illustrate the importance of tone, the word /ma/ means 'mother' in Tone 1, 'hemp' in Tone 2, 'horse' in Tone 3, and 'to scold' in Tone 4.

Table 1.5 Examples of the four Mandarin tones

Mandarin Tone	Description	Pitch	Examples in Pinyin	Gloss
1	high level	55	mā	mother
2	rising	35	má	hemp
3	falling-rising	214	mǎ	horse
4	falling	51	mà	scold

Neutral tone or Weak syllables

Another essential category of tone besides the four full or citation tones is neutral tone. Neutral tone has a short duration, and its vowel qualities are also reduced to short and lax schwa-like (Chao,1968). The duration of an unstressed syllable with neutral tone is 60% of the preceding stressed syllable (Cao, 1986); it shrinks to 50%- 60% of its stressed value (Lee, 2003; Chen and Xu, 2006). F0 space is also compressed (Li et al, 2014). Neutral tone is phonologically and morphologically conditioned; it usually occurs in a final unstressed syllable of a disyllabic word. 'Unlike the four basic tones, weak syllables cannot be pronounced in isolation; when an element which normally has weak stress is cited in isolation, ... it must be supplied with a [full] tone.' (Norman, 1988: 148).

There are four categories of neutral tone: the first three are rule-governed and restricted in number; the last one is lexically-based basis rather than rule-governed, i.e., it is irregular (see Lu, 1995; therefore the transcriptions of input speech in this thesis are based on mothers' real production). First, it regularly occurs in reduplication, the repeat of a syllable in a word. The reduplicated syllable loses its tone and takes on the neutral tone. For example, in kinship nouns: /ma1ma0/ 'mother' and /ti1ti0/ 'young brother' or in a noun or verb reduplicated to become a disyllabic word: /kou3kou0/ 'doggie', /ɕiŋ1ɕiŋ0/ 'star', /kʰan4kʰan0/ 'just look' and

/ɕiɛ4ɕiɛ0/ ‘thanks’. Second, morphemes that are clitic-like grammatical particles are always in neutral tone, for example, the word with aspect particle /lai2lə0/ ‘to come + aspect marker’ and the possessive /ni3tə0/ ‘your + possessive particle’. Third, it occurs on the affix of nouns normally /tsi/, for example, /pi2tsi0/ ‘nose’ and /ji3tsi0/ ‘chair’. Fourth, in two combined lexemes, especially nouns, the second lexeme is in neutral tone, for example, /ɚ3tuo0/ ‘ear’ and /t^{h}ou2fa0/ ‘hair’. Therefore, the neutral tone seems to occur no less often than the four basic tone categories in adult input.

There is a controversy concerning the nature of stress in Mandarin Chinese. In spite of the view that stress and non-stress in Mandarin are different from Western languages (non-tone languages), linguists agree that reduced duration and compressed pitch occur in unstressed syllables in Mandarin (Chao,1968; Shen, 1989; Lin, 2001a; Wang et al., 2002). In Wang et al. (2002:126)’s interpretation, in Mandarin every syllable which carries lexical tone can, in principle, be fully realized. The reason is that lexical tone needs sufficient syllable duration and a sufficiently wide pitch range to be duly produced and perceived. In certain words with a neutral tone, tonal syllables are de-stressed, and thus have their full realization reduced. “Stress (**重音**) in a word can only be established with regard and in opposition to non-stress (**轻音**); the stressed syllable does not have to be especially prominent (**重读**) at all.” We can interpret this idea in the disyllabic word with neutral tone /ji2kə0/ ‘one’. The first syllable is perceived as stressed, and the second syllable as unstressed. However, this impression is not due to /ji2/ (first syllable) being pronounced with extra amplitude. Rather, this can be attributed to the second syllable, /ke/, being unstressed (it has neutral tone, short duration and a possibly reduced or even deleted vowel). In fact, even if Tone 2 on the first syllable of /yi2ge0/ were realized in a subdued way, with a rather compressed pitch range and moderate duration, the word would still be perceived as a trochee (a strong-weak pattern).

In stress-timed languages like English, an unstressed syllable has shorter length, weaker prominence and reduced rime, e.g. /tə/ in ‘better’, /kən/ in ‘bacon’ (Lin, 2007). A neutral-toned syllable in Mandarin has all the characteristics of an unstressed syllable in a stress-timed language (Cheng, 1973; Lin, 2007); accordingly, we can treat these parallel structures as the same in phonological acquisition. That is, a disyllable with a full or citation tone and a neutral tone is treated as a trochaic word.

Tone sandhi

Tone sandhi refers to the phonological phenomenon in which a tone takes on the features of a neighboring tone under particular conditions. There are two such cases in Mandarin: first, in a sequence of two consecutive instances of Tone 3 (falling-rising contour), the first is changed to Tone 2 (rising), as shown in Table 1.6.

Table 1.6. Examples of Tone 3 sandhi rules. Sources from Sun (2006) and Lin (2007)

Pinyin	Gloss	Tone3 + Tone3		Tone2 + Tone3
shuǐ guǒ	fruit	214 214	→	35 214
hěn hǎo	very good	214 214	→	35 214

The second case is the tonal changes for two specific words, bu 'not' and yi 'one' (Table 1.7). The basic tone for bu is Tone 4, and it remains Tone 4 before Tone 1, 2 and 3, but changes to Tone 2 when followed by another Tone 4. The situation is a little complicated for the word yi 'one'. The basic tone for yi is Tone 1, which is only realised as a single syllable or at the end of a phase but changes to Tone 4 when it is followed by Tone 1, 2 or 3, and to Tone 2 when followed by Tone 4.

Table 1.7. Examples of bu and yi sandhi rules. Sources from Lin (2007)

Pinyin	Gloss	before T1	before T2	before T3	before T4
bù	not	[pu4xɤ1]	[pu4nɑn2]	[pu4xɑu3]	[pu2jɑu4]
yī	one	[ji4tiɛn1]	[ji4niɛn2]	[ji4wɑn3]	[ji2jɑŋ4]

It is necessary to discuss another phenomenon called reduction. The full tone value of the complex tone (Tone 3) is found only in sentence- or phrase-final positions; it is a bit shorter in phrase-initial position and does not preserve the full pitch contour. For instance, if Tone 3 is followed by another tone, it will be reduced by deleting the final rise and preserve only the low-falling part (Lin, 2007: 196): lǎoyīng [lɑu214 jəŋ55] → [lɑu21jəŋ55] 'eagle'. Therefore, the low falling tone (similar to Tone 4; there is no high/low distinction in same tone category in Mandarin, therefore we treat this as Tone 4 in our analyses) is more frequent

than the falling-rising tone (Tone 3) in Mandarin adult speech. Infants would hear a majority of falling contours in general.

1.5 Research questions and the contents of this thesis

The first goal of this thesis is to explore the nature of early ambient-language effects. An investigation of Mandarin-learning infants' babble vocalization is carried out. Because the 'babbling drift' concept is still under debate in phonological development research, we will present cross-linguistic evidence that justifies positing such a phenomenon. Our second goal is to explore Mandarin children's challenges in their segment and tone acquisition from a phonological point of view. Given children's immature and limited capacity in memory and articulatory control, how do they handle the challenge and how is their phonological system built and developed?

Accordingly, in this thesis we will explore four main research questions:

1. To what extent do lexical tone patterns such as those used in Mandarin occur in the pre-linguistic period? And to what extent do infants' tonal patterns in babble vocalizations reflect the phonology of the ambient language?
2. What characterises the phonology of Mandarin-learning children? What segmental advances and challenges will they have in their early word production? How is the phonology of the ambient language incorporated into their word production?
3. What tonal advances and challenges will they have in their early word production? How does the development of lexical tone production unfold?
4. Is there an influence of the specific adult input speech, for example, the frequency of particular categories to which different children are exposed in the same Mandarin community?

This thesis aims to fill the gap in research relating to the phonological development of Mandarin-learning children from an emergentist and holistic view. In Chapter 1 we have briefly reviewed the concept of an emergentist approach, the trajectory from babbling to the early word period and the role of input speech, and we have presented a sketch of Mandarin phonology. We will exemplify how the theory serves as a framework in Mandarin children's phonological development in Chapters 2, 3 and 4. In Chapter 2 we provide tone patterns in babble vocalizations of infants from two different languages. Drawing on Mandarin and

English data from 27 children (3 groups), we observe ambient-language effects in babbling. Before producing the first words, infants acquire some parts of their native language phonology. We discard the traditional method of language background judgements, arguing that that method disposes the results to reject the early language-specific phonological feature. We present an alternative analysis, using a quantitive method to explore the issue. In Chapter 3 we follow Mandarin-learning children over the single-word period, providing analyses of one session early and one session late in that period to trace their emergent production skills. We examine the characteristics of their prosodic structures, segment inventory and possible templates, illustrating how early accuracy is followed by later regression. We also analyse the data of children learning other languages, where information is available, to evaluate the similarities and differences. In Chapter 4 we move to the lexical tone development of the same Mandarin-learning children. Previous studies have focused on the order of acquisition of lexical tones and accuracy in production; here, we investigate tone learning as part of an emergent system. We examine 1) the occurrence, accuracy and substitution patterns of tones, (2) developmental trends in child production of each tone, and (3) possible factors accounting for the developmental patterns. Chapter 5 concludes the thesis with a summary of our main objectives, and interprets our findings in the light of models of phonological development.

1.6 Analyses

In the analyses throughout this thesis, all tones will be represented by numerals: Tone1 = high level tone, Tone 2 = rising tone, Tone 3 = falling rising tone and Tone 4 = falling tone. The expressions will be simplified further for the sake of convenience, especially in disyllables: T1T1= Tone 1+ Tone 1, T2T2= Tone 2+ Tone 2 and so on.

Unlike some other studies, which consider children's development by age, we distinguish children's level of advance based on lexical milestones. Two word-points are identified and applied in this thesis: the 4-word point (short for 4wp), the first month in which the child uses four or more identifiable adult-based words spontaneously in a half-hour session, and the 25-word point (25wp), the first month in which the child uses twenty-five or more identifiable words spontaneously. The four-word point is defined as the beginning of lexical use (Vihman, 1996). The twenty- five-word point is an important developmental point because it is the end of single word period and is considered as the time when some specific patterns can generally be observed (Renner, 2017). These two word-points correspond to a cumulative

lexicon of about 10 words and 50 – 70 words in the parental diary record, respectively (see Vihman & Miller 1988). Our comparison allows us to compare learning rates. For example, of 8 Mandarin-learning children recorded from 9 months to 18 months, four reached the 4-word point by 12 months; five had reached the 25-word point before the study ended.

At the 25-word point we distinguish between SELECTED and ADAPTED word productions. This is a comparison of each child word variant with its adult target form in relation to each prosodic structure. 'Selected' means child forms are approximately accurate or close enough to the target form, given the child's resources. If the prosodic structure of child's word matches its target regardless of segment or tone, it is treated as 'selected'. Word forms that are missing a structure that the child never produces could also be treated as selected, depending on each child's existing phonological knowledge in the session. For example, words with a missing coda are treated as selected for Mandarin-learning children who do not yet produce codas at all, but not for those who do produce them. 'Adapted' refers to child forms that are not a total match but are modified to a particular form that fits a given prosodic structure.

2. AMBIENT-LANGUAGE EFFECTS IN BABBLING: A STUDY OF TONE IN MANDARIN- AND ENGLISH-LEARNING INFANTS

2.1 Introduction

The vocal abilities of infants develop rapidly in the first year of life. As infants pass through a number of successive stages, the key milestone is the emergence of the production of speech-like syllables, so called 'babbling' (Oller, 2000). There are two incompatible theoretical perspectives on babbling: an 'independence' hypothesis that assumes that babbling is the natural output of physiological maturation and essentially independent of children's ambient language (e.g., Locke,1983; MacNeilage & Davis, 1993) and a 'babbling drift' hypothesis. Brown (1958) first proposed 'babbling drift' to refer to the idea that an infant's babble vocalizations gradually shift in the direction of the ambient language, even before identifiable words are observed.

It is likely that auditory exposure to adult speech plays a key role in vocal development. It has been shown that deaf infants have very late onset of canonical babbling, after the first year (Oller & Eilers, 1988; Oller, Eilers, Bull, & Carney, 1985). Also, different from typically developing infants, hearing-impaired infants have a limited phonetic repertoire (Moeller et al., 2007). Auditory experience seems to help infants shape babbling and to facilitate phonological learning through social interaction in their first year (Gros-Louis et al., 2006).

Prosodic features could be the first to display influence from the ambient language. Infants show perceptual sensitivities to some specific rhythmic properties of language from birth, as evidenced by neonates' preference for listening to sentences of their native language as opposed to sentences of a rhythmically different language (Moon et al., 1993; Mehler et al., 1988). From a production perspective, in a prosodic-feature-oriented discrimination task, French-speaking adults were asked to make perceptual judgements discriminating 8- to 10-month-old French-learning infants' babble vocalizations from those of Arabic- and Cantonese-learning infants (Boysson-Bardies, Sagart & Durand, 1984). This was found not to be a hard task. In a follow-up study, three groups of French speakers were asked to listen to French and Arabic infants' vocalizations at 6, 8 and 10 months. Although they failed with the 10-month-old sample, they performed well in the two younger groups. The investigators interpreted this as being due to the utterances from the 10-month-old samples being "poorer

in prosodic cues" (p.10). They concluded that the successful discrimination was based on prosodic features – for example, weak-strong rhythmical patterns were prominent only in the Arabic infants' babbling.

Ambient language effects have been revealed by acoustic analyses of the vowel space. Boysson-Bardies et al. (1989) sampled the babbling of five 10-month-old infants exposed to British English, French, Arabic and Cantonese. Their results showed that English infants produced more front vowels, French infants produced more mid central vowels, and Cantonese infants produced more low central vowels. The same favored vowel patterns were found in adults' running speech. For example, English favors high, front vowels (/i/, /e/), whereas Cantonese favors low back vowels (/ɑ:/, / ɔ:/). This indicates that those infants 'set loose articulatory limits to tongue and lip movements as a first step toward acquiring the vowel system of the ambient language' (Boysson-Bardies et al., 1989, p.14).

Similarly, for consonant production in babbling, Boysson-Bardies and Vihman (1991) investigated place and manner of articulation of consonants produced by English, French, Japanese and Swedish infants from 9 or 10 to 17 months. A larger proportion of labials were found in the vocalizations of 10-month-olds exposed to English and French than in those of infants exposed to Japanese or Swedish, which reflects the distribution of these consonant types in the adult language (see Vihman et al.,1994, for the adult languages).

Ambient language effects on intonation in babbling have also been demonstrated in cross-linguistic studies. Six- to twelve-month-old English- and French-acquiring infants' di- and trisyllabic babble vocalizations were analysed in Whalen, Levitt & Wang (1991). English infants' utterances were dominated by falling pitch contours, while the French infants produced significantly more rising contours than their English peers. This pattern is consistent with the main difference in pitch contours in the two adult languages (cf., e.g., Delattre, 1961).

The babbling drift or ambient language effect continues to be a subject of controversy. Thevenin et al. (1985) failed to observe specific ambient-language effects in a listening task based on the vocalizations of 7- to 10- and 11- to 14-month-old English-speaking and Spanish-speaking infants. A number of studies have examined the segmental repertoire in the babbling of children acquiring different languages. Both Spanish- and English-learning

infants were found to produce predominantly CV syllables with voiceless, unaspirated plosive consonants and similar proportions of various vowel types (Oller and Eilers, 1982). Vihman (1992) explored the early syllables of 33 children aged 9 to 16 months in four languages (English, French, Swedish, Japanese). The data showed that syllables involving [d] and [h] and vowels including [a/æ] and [ə/ʌ] are most commonly used by all the language groups. Similarly, both Korean- and English-learning infants showed a preference for stop and labial consonants compared to other consonants (Lee, Davis & MacNeilage, 2010). This has been explained by a biological model of perceptual and motor factors (Locke, 1983; Kent, 1992).

In the literature on the extent of language-specific effects on babble, segmental and prosodic aspects of the productions of infants of different language backgrounds have been presented. However, few attempts have been made to explore infants' babble vocalizations in relation to lexical tones. Therefore, it is not clear when tone-learning infants start their tonal category learning. How is this aspect of the phonology of the ambient language incorporated into babbling?

In tonal languages like Mandarin Chinese, pitch contour is used to minimally encode lexical differences. Lexical tones typically reflect variations in fundamental frequency within a single syllable (e.g., Burnham & Mattock, 2007; Liu & Samuel, 2004). In non-tone languages like English, pitch also conveys information and varies in meaningful ways, communicating pragmatic and emotional meaning (Bolinger, 1989; Hirschberg & Ward, 1992; Pierrehumbert, 1980; Ward & Hirschberg, 1985), highlighting aspects of grammatical structure (Gussenhoven, 2004) and differentiating talker identity, gender, and register (Fernald, 1992), but it is only used for longer units as part of intonation patterns, not in a lexically distinctive way. English-learning children in Quam and Swingley (2010) were unable to exploit contrastive pitch contours when learning new words, reflecting the fact that pitch contour variation is not lexically contrastive in English.

First, this chapter aims to examine the ambient-language effect in babbling, focusing on the use of tone in Mandarin- and English-learning infants. A recent study investigating ambient-language effects includes Mandarin-acquiring infants, but it failed to find evidence of such effects on their babbling utterances (Lee et al., 2017). Inspired by Engstrand, Williams, and Lacerda (2003), who chose vocalizations of American and Swedish infants at 12 months and

18 months, Lee et al. looked at Mandarin- and English-acquiring infants' vocalization at 8, 10 and 12 months and found ambient-language effects only in the period of emerging words (12 months), not earlier, in 'pure babble'. However, we argue that the method of forced-choice listening to language background used by Lee et al. (2017) may tend to mask early language-specific effects. We will apply a combination of methods to examine the babble vocalizations of Mandarin- and English-learning infants quantitatively and qualitatively.

Second, this chapter investigates the early phonological characteristics in terms of tone in Mandarin-learning infants. Only one study to date has investigated Mandarin infants' vocalizations in the pre-linguistic period. Chen and Kent (2009) examined tonal patterns in 24 Mandarin infants' vocalizations, from the canonical babbling period to the point of having produced their first fifty words. Children were assigned to a young group (0;7 to 1;0) and an older group (1;1 to 1;6). The results showed a prominent use of falling tones (Tone 4) in the young group and use of the level tone (Tone 1) was low in the young group but higher in the older group. Although an adequate number of infants were examined in the pre-linguistic period, children in this study contributed only a single recording each and the sampling was not evenly spread across age points. Thus, it is unknown how frequently and consistently individual children produce each of the tone patterns in babbling. We will analyse frequencies of occurrence of tone categories in Mandarin and English infants' babble vocalizations over four months. Moreover, all speech-like utterances, including disyllabic and longer vocalizations, were selected for pitch comparison in Chen and Kent (as in Lee et al., 2017); no attention was paid to the syllable level. However, lexical tone in Mandarin occurs on every single syllable. Therefore, we will separately analyse tone patterns in infants' vocalizations in monosyllables and in disyllables.

Mandarin and English have distinct pitch/tone characteristics. These differences can facilitate consideration of the emergence of ambient language patterns in babbling. In this chapter, the production of tonal patterns by infants exposed to Mandarin will be compared with that of infants learning English to address the following questions: 1) Is tone detectable in both Mandarin- and English-learning infants' babble vocalizations? If so, to what extent are the tones identified in babbling the same or different in Mandarin- and English-acquiring infants? 2) To what extent does the target language exert an influence on infant use of such tonal patterns? 3) Does a caregiver's speech directly influence her Mandarin-learning infant?

The ‘babbling drift’ hypothesis holds that phonetic characteristics of the ambient language appear from an early age. If intonation is a prime phonetic carrier of emotion, tonal features should attract infants’ attention and appear early in babbling. Therefore our hypothesis was that children exposed to Mandarin, which has lexical tone, will tend to produce more identifiable tones in their babble than children exposed to English, which does not.

2.2 Methods

Participants

Twenty-seven monolingual infants served as participants, divided into three groups. The first group comprised nine Mandarin-learning children (six girls and three boys) who were born and raised in North China (Beijing and Shandong). The second group consisted of eight Mandarin-learning children (five girls and three boys) who were recruited in York, from families that speak only Mandarin as their home language. The third group consisted of eight English-learning infants (three girls and five boys) who were randomly chosen from an existing database of children from North Yorkshire, England (the University of York corpus; see DePaolis et al., 2016). None of the participants had any known developmental or hearing problems.

Recording Procedure and Data Collection

Infants were recorded at home in naturalistic play interactions with their caregivers from age 9 to 12 months. Half-hour recording sessions were conducted on a biweekly basis or as frequently as possible. The infants were asked to wear a vest which has a microphone in a hidden pocket and the mothers had a microphone fitted to their clothing. For York Mandarin- and English-learning infants, the recordings were made using a Sony digital video camera recorder, either HDV 1080i HVR-A1E or DSR-PDX10P on 2 channels using 2 Sennheiser Camera Set wireless microphones ew112-p G2 microphones, one worn by the mother and one worn by the infant. Mandarin-learning infants in North China were audio-recorded only, by iPhone, with hidden wired microphones. This study was approved by the Ethics Committee of the University of York. Written informed consent was obtained from all participating families.

Transcription

The recordings were transferred digitally to a computer. A native Mandarin speaker (the author of this thesis) listened to all the utterances and transcribed phonetically using ELAN Linguistic Annotator.

Data analysis

This analysis aims to assess the similarity and differences in each child's use of tone across four months in the pre-linguistic period. In order to analyse the frequency of the tonal patterns, all babble vocalizations were transcribed to obtain a representative sample for each child. Exceptionally long vowels, cry, fussing, laughter, creaky sounds and overlap with noise or other speech were excluded. Infants' tone-bearing vocalizations were perceptually transcribed and broadly classified in terms of the four basic tones of the adult Mandarin system. These potentially tone-bearing vocalizations were grouped into units by length in syllables, e.g. monosyllable, disyllable. Vocalizations composed of two shorter sequences with intervals not exceeding 2 secs were considered to be disyllables. We limited our attention to mono- and disyllables here for two reasons. First, these represented the predominant category in the child productions across the two languages. One- and two-syllable words are the majority (94.14%) of word types both in the Mandarin CHILDES database (Wang et al., 2010) and in English infants' production (Demuth & Fee 1995, Vihman & Miller 1988). Second, unlike longer multisyllables, these syllables permit comparatively clear analysis of pitch.

All recorded sessions were transcribed, but the session with the most vocalizations was selected for each month. Infant productions of each tone were tallied for each session. Unlike other studies reporting ambient-language effects based on a single observation period of infant's babbling, we make use of criteria like those applied to arrive at vocal motor schemes (McCune & Vihman, 2001), based on frequency and longitudinal stability, to evaluate the occurrence of tones in children's early vocalizations – especially for disyllables, which are not analysed statistically here. Specifically, in disyllables a tonal pattern produced *a minimum of 20 times in two or three out of four contiguous sessions* was taken to be a frequent tonal pattern.

Background information on Mandarin phonology is provided in Chapter 1. In this chapter, all recognisable pitch patterns will be represented by tone with numerals. All English-learning

infants' pitch patterns are labelled as tones here to permit direct comparison, using the categories of Mandarin tones. In order to classify all pitch patterns as clear adult Mandarin categories, if we found complex patterns within one syllable (which can be seen as a combination of two tones on a single syllable (e.g., Tone1-Tone4, Tone2-Tone4 or Tone4-Tone2, a phenomenon which does not occur in adult Mandarin), such cases are treated as *unidentifiable*.

2.3 Descriptive analysis and results in three groups

The analyses will begin with data from the Mandarin-learning infants in North China, which will give us a first view of what Mandarin-acquiring infants' tone-bearing babble vocalizations look like. That will be followed by the analysis of York Mandarin-acquiring infants' babble vocalizations, to illustrate possible variation in the use of tone in the same language. Then an analysis of English infants' babble vocalizations is presented, to show children's babbling patterns in two different language groups and to identify possible similarities and differences in the types of tones that can be found in the babble of children learning different languages.

Group 1 - Mandarin-acquiring infants in North China

These analyses present the North China Mandarin-learning infants' tonal patterns child by child and month by month, first in monosyllables, then in disyllables, and finally in terms of their distribution by length in syllables across the four months. The use of tone within monosyllables in the North China Mandarin-learning children, from 9 to 12months, is shown in Table 2.1. Criterion for monosyllabic use was set to *over 10 uses in a session.*

Table 2.1. Occurrence of tones in North China Mandarin-acquiring children's monosyllabic vocalizations (bold face indicates over 10 productions in a session)

Name		Tone1	Tone2	Tone3	Tone4	Total	uncat
DD	0;9.27	**10**	3	0	**11**	24	
	0;10.16	**12**	3	0	**13**	29	
	0;11.6	**18**	5	0	**20**	43	
	1;0.3	8	6	0	**15**	29	

JD	0;9.29	**12**	5	0	**10**	27
	0;10.17	**24**	4	0	**21**	49
	0;11.13	**14**	4	0	**14**	32
	1;0.6	**11**	4	0	**10**	25
TZ	0;9.6	**35**	3	0	**16**	54
	0;10.3	**15**	6	0	**25**	46
	0;11.26	**12**	3	0	**17**	32
	1;0.8	**23**	3	0	**38**	64
WZ	0;9.7	6	4	0	9	19
	0;10.4	8	4	0	4	16
	0;11.6	**15**	4	0	**10**	29
	1;0.29	6	3	0	**14**	23
SL	0;9.26	**10**	**20**	0	**23**	53
	0;10.22	9	**11**	1	**15**	36
	0;11.12	8	**13**	1	**18**	40
	1;0.13	9	**15**	0	**21**	45
KL	0;9.7	**37**	**10**	0	**25**	72
	0;10.12	**21**	**12**	1	**21**	55
	0;11.9	**24**	**11**	0	**13**	48
	1;0.14	**19**	9	0	**13**	41
LL	0;9.21	9	6	0	**10**	25
	0;10.20	**27**	**15**	0	**31**	73
	0;11.6	**10**	5	0	**12**	27
	1;0.13	**11**	5	0	**13**	29

WW	0;9.7	7	3	0	**12**	22
	0;10.23	**14**	7	1	**15**	37
	0;11.2	**12**	9	0	**13**	34
	1;0.9	**10**	6	0	**12**	28
QQ	0;9.14	**17**	**12**	0	**23**	52
	0;10.26	**22**	**19**	0	**26**	67
	0;11.18	**12**	9	0	**25**	46
	1;0.9	**10**	**12**	0	**31**	53

Table 2.1 provides an overview of monosyllabic tonal patterns in the North China Mandarin-learning infants' vocalizations: Tone 4 occurred the most often across all sessions; Tone 1 was also produced frequently and Tone 2 less often. Only a few possible instances of Tone 3 were identified in any of the children. Therefore, in general, these Mandarin infants all produced a majority of Tone 1 and Tone 4 within monosyllables. Tone 2 occurs to criterion in three children's monosyllabic vocalizations.

If we look closely at the use of each tone in each child, although most of these children had access to Tone 1 and Tone 4, they differed in their proportions of use and trajectories. Three infants (DD, JD and WW) consistently produced similar amounts of Tone 1 and Tone 4, above the criterial level of 10 uses. Four infants (TZ, SL, LL and QQ) produced more Tone 4 than Tone 1 in at least three consecutive sessions. One infant (KL) produced more Tone 1 than Tone 4 across all sessions. One child had no consistent preference, producing more Tone 1 in some sessions and more Tone 4 in other sessions.

Notice that unlike the others, three children preferred Tone 2 consistently throughout the babbling period covered here. SL used large numbers of Tone 2 in all four of her monthly sessions. Tone 2 rather than Tone 1 (which was above criterion only at 9 months) played as dominant a role as Tone 4 in her monosyllabic vocalizations. Two other infants (KL and QQ) also heavily produced Tone 2, but often produced Tone 1 and Tone 4 as well.

Now we move to tones in North China Mandarin-learning infant's disyllabic vocalizations. The infants produced far fewer disyllables overall than monosyllables. Therefore, the criteria had to be adjusted accordingly, to *20 occurrences in total across four months and occurrence in at least two consecutive sessions*. Unlike the monosyllables, there is considerable diversity in the infants' disyllabic tonal patterns (see Appendix I). Following the adjusted criteria, there are some tendencies for most infants: combinations of Tone 1-Tone 4 and Tone 2-Tone 4 were the most frequent patterns within disyllables, both in terms of tokens within child and in terms of the number of children producing them. The Tone 1-Tone1 association was the next most common pattern. Other combinations are produced only rarely and by few children or inconsistently.

When we compare disyllabic tonal patterns across children we see that the major tonal patterns that the children preferred vary by child (Table 2.2). Four children have Tone 1-Tone 4 as their characteristic pattern. Three children used Tone 2-Tone 4 most. Two children produced Tone 1-Tone 1 most. Two children had not yet built a disyllabic tone scheme.

Table 2.2. Summary of most used patterns for each child

Child name	Critical patterns
TZ	Tone 1-Tone 4
WZ	Tone 1-Tone 4
QQ	Tone 1-Tone 4
LL	Tone 1-Tone 4; Tone 2-Tone 4
DD	Tone 2-Tone 4
KL	Tone 2-Tone 4; Tone 1-Tone 1
SL	Tone 1-Tone 1
JD	unclear
WW	unclear

Now we consider the length in syllables children produced each month during the babbling period.

Table 2.3. Distribution of tonal-bearing monosyllables and disyllables in China Mandarin children's vocalizations (**bold face** indicates the major pattern)

Child Name	9 months		10 months		11 months		12 months	
	Mono	Di	Mono	Di	Mono	Di	Mono	Di
DD	**53%**	47%	**71%**	29%	**58%**	42%	**55%**	45%
	24/45	21/45	**29/41**	12/41	**43/74**	31/74	**29/53**	24/53
JD	**87%**	13%	**84%**	16%	**80%**	20%	**74%**	26%
	27/31	4/31	**49/58**	9/58	**32/40**	8/40	**25/34**	9/34
TZ	**86%**	14%	**79%**	21%	**57%**	43%	**66%**	34%
	54/63	9/63	**46/58**	12/58	**32/56**	24/56	**64/97**	33/97
WZ	33%	**67%**	39%	**61%**	48%	**52%**	46%	**54%**
	11/33	**22/33**	16/41	**25/41**	29/60	**31/60**	23/50	**27/50**
SL	**61%**	39%	**55%**	45%	**55%**	45%	**60%**	40%
	53/87	34/87	**36/66**	30/66	**40/73**	33/73	**45/75**	30/75
KL	**72%**	28%	**66%**	34%	**56%**	44%	**59%**	41%
	72/100	28/100	**55/83**	28/83	**48/85**	37/85	**41/69**	28/69
LL	**52%**	**48%**	**51%**	**49%**	**50%**	**50%**	**52%**	**48%**
	25/48	**23/48**	**73/142**	**69/142**	**27/54**	**27/54**	**29/56**	**27/56**
WW	**61%**	39%	**79%**	21%	**79%**	21%	**53%**	47%

	22/36	14/36	**37/47**	10/47	**34/43**	9/43	**28/53**	25/53
QQ	**81%**	19%	**80%**	20%	**77%**	23%	**72%**	28%
	52/64	12/64	**67/84**	17/84	**46/60**	14/60	**53/74**	21/74

Here we see a categorical similarity across the children (Table 2.3): The majority had a strong preference for monosyllables. Seven out of nine children produced primarily monosyllables over all four months investigated here. However, one child (WZ) preferred disyllables. One child (LL) had equally balanced numbers of monosyllables and disyllables.

Group 2 - York Mandarin-acquiring infants

Now we move to another Mandarin group, in which the infants were raised in York (one was born in North China and moved to York when she was 8 months old), but exposed only to Mandarin during their babbling period. The tonal patterns in the York Mandarin-acquiring infants' monosyllabic vocalization are shown in Table 2.4.

Table 2.4 Occurrence of tones in York Mandarin-acquiring children's monosyllabic utterance (**bold face** indicates over 10 productions)

		Tone1	Tone2	Tone3	Tone4	Total	Uncat.
ZuoZuo	0;9.	**16**	**10**	2	**20**	48	2
	0;10.	**19**	**20**	0	**25**	64	3
	0;11.	**12**	**10**	0	**13**	35	1
	1;0.	**10**	**11**	0	**18**	39	1
Weilun	0;9.	8	0	0	**18**	26	2
	0;10.	**14**	7	0	**18**	39	3
	0;11.	6	3	0	**16**	25	0
	1;0.	**10**	1	0	**15**	26	0

Keke	0;9.	3	1	0	6	10	0
	0;10.	**13**	0	0	**15**	28	0
	0;11.	**13**	3	0	**13**	29	0
	1;0.	9	5	0	9	23	0
Yifan	0;9.	**11**	**11**	0	**11**	33	2
	0;10.	7	7	0	**12**	26	2
	0;11.	5	3	0	**10**	18	0
	1;0.	4	2	0	**13**	19	0
Didi	0;9.	5	2	0	**11**	18	0
	0;10.	8	2	0	**15**	25	1
	0;11.	9	3	0	**20**	32	1
	1;0.	5	6	0	**18**	29	0
Xinyu	0;9.	**40**	4	0	**69**	113	0
	0;10.	**10**	4	0	**16**	30	0
	0;11.	**10**	4	0	**18**	32	0
	1;0.	**10**	6	0	**17**	33	0
Yiyi	0;9.	9	8	1	2	20	0
	0;10.	**16**	4	0	**12**	32	0
	0;11.	**14**	3	0	**20**	37	0
	1;0.	**13**	0	0	**20**	33	0
Shi	0;9.	**29**	4	0	**11**	44	0
	0;10.	**10**	1	0	7	18	0
	0;11.	6	3	0	8	17	0
	1;0.	9	2	0	**13**	24	0

We can see that Tone 1 and Tone 4 are the most common patterns in these infants as well; Tone 4 is more frequent than Tone 1 overall in terms of the number of sessions in which it occurs. Tone 2 occurs much more rarely in monosyllabic vocalizations.

Almost all the children (7 out of 8) produce Tone 4 above criterion in monosyllables for at least two consecutive sessions. Most children (5 out of 8) have Tone 1 in their repertoire. Only one child production of Tone 2 meets the criteria in addition to his Tone 1 and Tone 4.

Overall, North China Mandarin-learning and York Mandarin-learning infants showed similar tendencies. They all produced Tone 4 the most frequently, Tone 1second most frequently and Tone 2 the least.

Group 3 - York English-acquiring infants

To see if an ambient language effect of tone use can be identified, English-learning infants' babble vocalizations are analysed in the same way as the Mandarin infants' babble. We start from tonal patterns in their monosyllabic vocalizations.

Table 2.5. Occurrence of tones in York English-acquiring children's monosyllabic utterance (**bold face** indicates over 10 productions)

		Tone1	Tone2	Tone3	Tone4	Total	Uncat.
AC	0;9.	**15**	7	0	**37**	59	3
	0;10.	5	9	0	**17**	31	5
	0;11.	0	**10**	0	**15**	25	4
	1;0.	4	7	0	**20**	31	5
AH	0;9.	3	2	0	**13**	18	2
	0;10.	1	2	0	3	6	0
	0;11.	0	1	0	5	6	0
	1;0.	2	4	0	**36**	42	3
CL	0;9.	1	2	0	**18**	21	1

	0;10.	5	4	0	**11**	20	3
	0;11.	1	2	0	**12**	15	3
	1;0.	9	3	0	**24**	36	2
CS	0;9.	7	3	0	**11**	21	2
	0;10.	2	1	0	**12**	15	1
	0;11.	1	0	0	**23**	24	4
	1;0.	2	0	0	**15**	17	2
GR	0;9.	1	**10**	0	**20**	31	2
	0;10.	0	1	0	6	7	1
	0;11.	0	4	0	**10**	14	0
	1;0.	2	1	0	**29**	32	5
JY	0;9.	2	1	0	**10**	13	1
	0;10.	1	1	0	**11**	13	1
	0;11.	2	3	0	**15**	20	4
	1;0.	6	**13**	0	**21**	40	6
RT	0;9.	3	9	0	**29**	41	1
	0;10.	1	0	0	**11**	12	0
	0;11.	0	3	0	**11**	14	2
	1;0.	7	**10**	0	**25**	42	3
S	0;9.	3	8	0	**16**	27	
	0;10.	2	4	0	**19**	25	
	0;11.	2	0	0	**12**	14	
	1;0.	0	3	0	**22**	25	

It can be seen from Table 2.5 that Tone 4 made up the largest number overall; other tone categories accounted for only about a third of English-learning infants' monosyllabic vocalizations. Relatively infrequent instances of Tone 1 and Tone 2 were observed. Therefore, it can be concluded that unlike the Mandarin-learning infants, who produced both Tone 1 and Tone 4 the most, the English-learning infants all produced a majority of Tone 4 within monosyllables.

There are also variations across the English infants. Four infants produced Tone 2 more than 10 times, but all in a single session. One child produced Tone 1 over criterion only in her 9-months session.

Note that, unlike the Mandarin-learning infants, most of whose monosyllabic utterances had simple patterns – either Tone 1, Tone 2 or Tone 4, English infants not only had Tone 4 as the most frequent, but also more uncategorized patterns, which means that more than a single pitch occurred in the one syllable.

We move to English-learning infants' tonal patterns in disyllabic vocalizations. English-learning infants produced fewer types of disyllabic tones than the North China Mandarin-learning infants did, therefore, only the occurring patterns are listed in Appendix II. Here we can see that Tone 2-Tone 4 and Tone 4-Tone 4 are the most frequent patterns across all the English infants. Other patterns are produced either rarely or by few children. However, apart from Tone 2-Tone 4, Tone 1-Tone 4 and Tone 1-Tone 1 characterize Mandarin-learning infants' patterns; Tone 4-Tone 4 is not detected as a preferred pattern in Mandarin-learning infants. It can be concluded that in disyllables, English-learning infants tend to use Tone 4-related patterns while Mandarin-learning infants tend to use both Tone 1- and Tone 4-related patterns.

To summarise, lexical tones used in adult Mandarin were detectible in both monosyllables and disyllables of Mandarin-learning infants' babble vocalizations. The different tonal patterns were observed in both monosyllables and disyllables in the two language groups. For English-learning infants, the falling contour (Tone 4) reflects only global intonation rather than lexical tone, while this contour also plays a significant lexical role in Mandarin. Mandarin-learning infants start to show adult tonal patterns, although their realizations of Tone 1 and Tone 2 are still infrequent. The results support the conclusion that infant babbling

is not only based on patterns common to both of these very different ambient languages. Children's experience with the specific language of the environment could also be seen to have an effect in the pre-linguistic period.

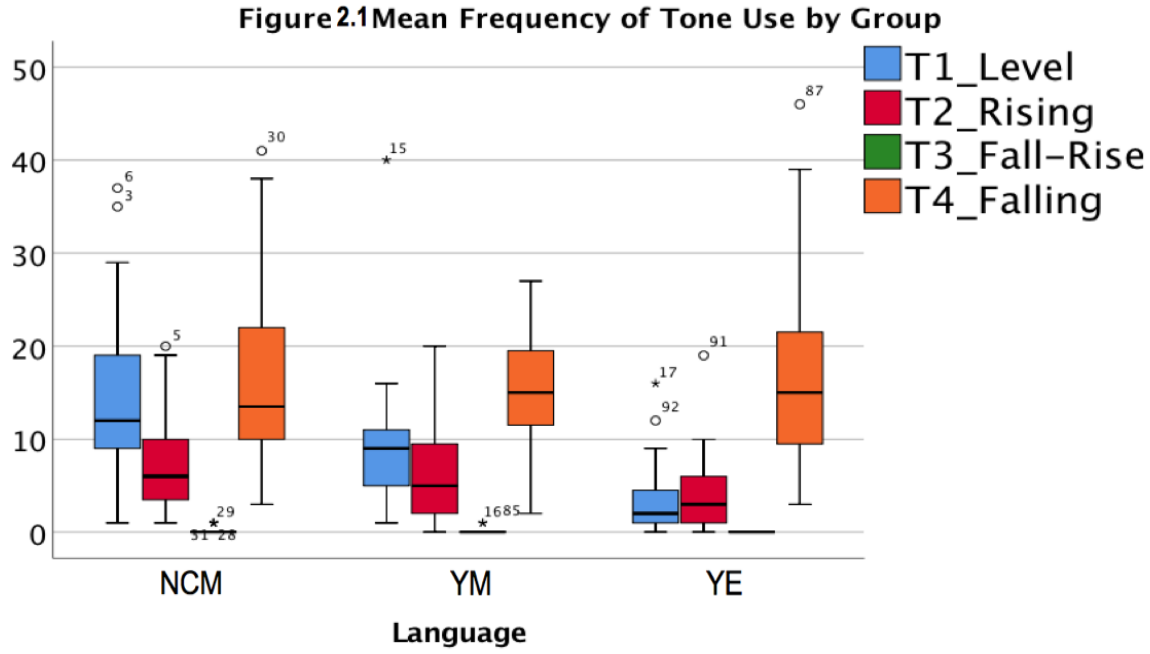
2.4 Statistical analyses and results in three groups

Differences were detected in the frequency of tone occurrence in babble vocalizations of infants exposed to different languages (tone-language vs non-tone-language) in our descriptive analyses, but statistical analyses are needed to test for the significance of the differences. Intergroup comparisons were conducted to test the hypothesis that linguistic background influences the distribution of tone categories in infants' pre-linguistic period. Differences between the North China Mandarin group (NCM hereafter), York Mandarin group (YM) and York English group (YE) were analysed with the nonparametric Mann–Whitney U-test and Kruskal–Wallis test. All analyses were carried out using SPSS Statistics, version 25.

Counts of tone patterns in monosyllables were made for each subject at each monthly session. The overall productions of tone-bearing monosyllables throughout the 4-month period studied within each group are displayed in Figure 2.1. The distributions of falling Tone 4 are consistently the most frequent across all three groups from 9 to 12 months. High-level Tone 1 was produced more frequently than rising Tone 2 by Mandarin children and at the same low frequency by the English children. Tone 3 occurred rarely (and in Mandarin-learning children only).

There is a significant difference between language groups in the use of Tone 1 . Pairwise comparison shows that both North China Mandarin infants and York Mandarin infants used significantly larger proportion of level pitch (Tone 1) than the English group ($p < .001$ [est. = 1.31], and $< .001$ [est. = 0.81], North China and York Mandarin samples, resp.). There is also a small difference between the North China group and the York Mandarin group ($p=.045$) due to the higher percentage of Tone 1 produced by the North China group. Analysis of variance by Tone 2 shows the significant difference between the North China group and the York English group ($p<.001$). The York Mandarin group does not differ from either the North China group or the York English group. There is no significant different between the three groups in the distribution of Tone 4. The interaction between tone distribution and age is not significant for any tone ($p> .01$).

Figure 2.1 Mean frequency of tone use by group



Conclusion

Ambient-language effects can be found in the babble vocalizations of children in the three different language environments. The Mandarin-learning infants in both groups produced significantly more Tone 1 than the English infants. And the preference for Tone 1 was also consistently found over the four monthly sessions in both Mandarin groups. This confirms the finding from previous studies, that level contour was the most or the second-most common pattern (Kent & Murray, 1982; Robb et al., 1989). In addition, the greater use of Tone 1 in the infants living in a monolingual community in North China further suggests the same ambient language effect. Regarding disyllables, Tone 1+Tone 1 and Tone 1+Tone 4 combinations are the most frequently occurring patterns in Mandarin-learning infants, but not in English infants, who produce Tone 2+Tone 4 and Tone 4+Tone 4 the most. The Tone 1-related patterns are shown in children learning Mandarin only. It seems that Mandarin children's disyllabic tone-bearing vocalizations tend to reflect their language environment.

Universal characteristics of tone patterns, which may result from general articulatory constraints – that is, control of the rate of vocal cord vibration – were also observed. English-acquiring infants produced predominantly Tone 4 (falling contours) in monosyllables and Tone2-Tone4 combinations (rising and falling contours) within disyllabic utterances. This tendency is in accord with the findings of previous studies (Kent & Bauer, 1985; Robb et al.,

1989; Whalen, Levitt &Wang, 1991) and the adult language. In spite of gross lexical tone differences between Mandarin and English adult languages, Mandarin infants also produced large numbers of Tone 4 within monosyllables and Tone 2-Tone 4 as one of the most frequent patterns in disyllables. The falling contour in pitch reflects a natural drop in subglottal pressure over the course of an utterance, which is effortless for infants from an early age.

Moreover, ambient language effects are present in the various pitch changes within syllables. In this study the mono- and disyllabic utterances of Mandarin children involved less complex patterns (rise-fall, fall-rise) within one syllable than those of English children. In terms of prosodic patterns, English and Mandarin differ as to whether f0 changes signal a change in meaning. As Mandarin is a tone language, tones characterize every single syllable; therefore, children may be learning to attend to f0 changes at the syllable level. On the other hand, for a non-tone language f0 correlates only with intonation, which applies to phrasal structures; thus children do not need to learn to keep the pitch value steady at the syllable level.

2.5 Experiment 1: Adult Perception Test

There is no reliable automated acoustic software for determining tone, which requires not only identifying the fundamental frequency, but also characterizing its trajectory as belonging to one of the four tones. Perceptual or acoustic analysis to assess ambient-language effects in the pre-linguistic period has been used in previous research. Our study uses both perceptual analysis, with adult listeners classifying the tone patterns, and instrumental analysis, comparing two spectrograms based on F_0 analysis. We chose a combination of methods to calculate the reliability of the tone judgments made by the researcher for a practical reason: It makes it possible to assess a large number of utterances with a group of examiners.

2.5.1 Method

Studies of the early appearance of ambient-language effects have involved adult perception of differences in the vocalizations of children from different language backgrounds. To investigate the reliability of phonetic transcription in the present study, an adult perceptual test was conducted. Instead of using traditional forced-choice language-background judgments, which have often failed to provide evidence of ambient-language effects (Thevenin, et al, 1985; Lee et al., 2017), our approach required adult listeners to judge the

tone patterns of each child vocalization based on the Mandarin category labels (Tone 1, Tone 2, Tone 3 or Tone 4).

Given concerns as to the quality of recording in the North China Mandarin group (on iPhone), only the York Mandarin and York English groups were chosen for the experiment. Children in the two York groups were recorded in the same location and with similar equipment. Ten utterances from 7 each of York Mandarin- and York English-learning infants were randomly extracted from the tone-bearing monosyllables in their 12-month recording session, with care to ensure that (1) no other voices could be heard, (2) there was no audio distortion and (3) audibility was comparable. There were 140 utterances in total, half from English and half from Mandarin recordings. The York Mandarin- and York English-learning infant data from which the utterances were drawn were collected independently; therefore, before presenting utterances to listeners for the present study, we normalized the sounds so that they were at the same mean amplitude level and no differentiating sounds could be heard.

Forty native speakers of Mandarin participated in the listening task (20 males, 20 females; mean age 25.7). All the Mandarin listeners are students at the University of York with English being their second language. None of them had a history of hearing loss.

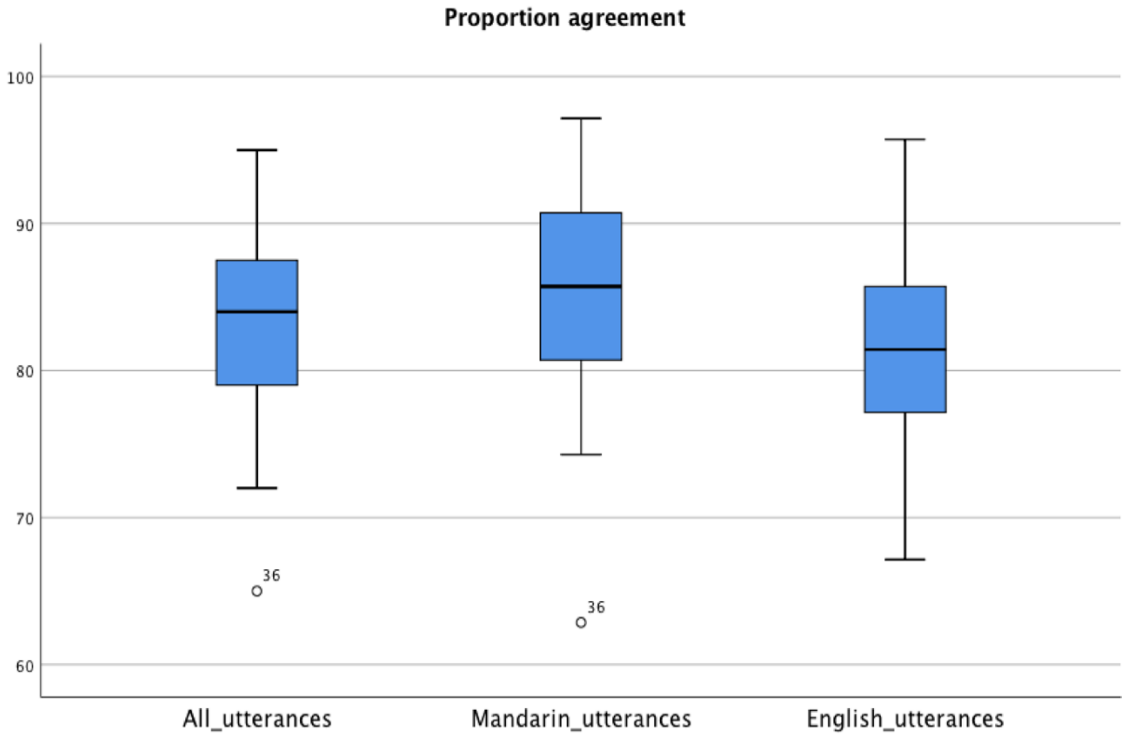
The listeners were presented through headphones with 140 babbling productions in random order. The listeners were unaware of the language environment of the babblers. For each utterance, listeners were given four choices on the center of the laptop screen: Tone1, Tone2, Tone3 or Tone4. After the decision was made, listeners were asked to press the next button to continue. Each utterance was played twice automatically, but listeners were allowed to listen as many times as they wanted.

2.5.2 Results

The responses of listeners were binary-coded, 0 = same as transcription (agree) and 1 = different from transcription (disagree). For each utterance, the proportion of agreement was calculated. We present the results and figures using proportion agreement. The mean overall agreement of the listeners for the entire set of utterances was 83%, 85% for Mandarin utterances and 81% for English utterances. Figure 1 shows the listeners' agreement across the whole set of utterances from the Mandarin- and English-acquiring infants, and for the

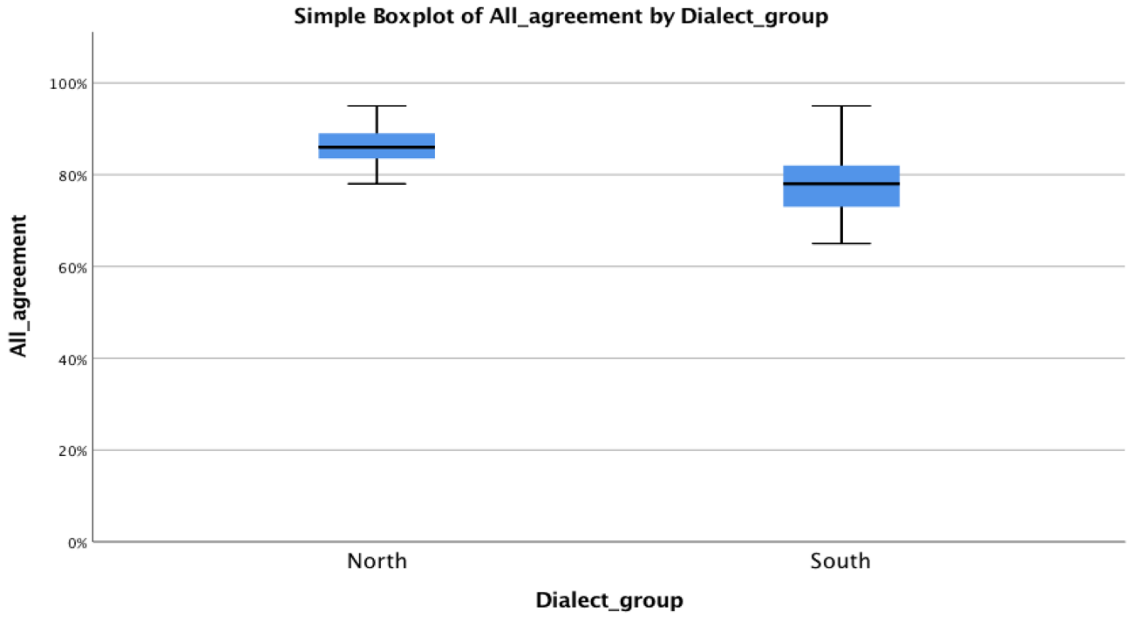
utterances of the two language groups separately. Listeners showed better than chance agreement overall with no significant difference in performance on Mandarin vs. English babble.

Figure 2.2 Listener's agreement of tone patterns in infants' babble vocalizations.



Mandarin is the official language of China, but a range of different dialects are also widely spoken. Although all dialects have tone, they can be broadly divided into two main branches, a northern Mandarin-based branch and a southern branch with more lexical tone categories. To see if the listeners' dialect background affected their performance, we analyse their scores on a dialect basis. Given that the researcher's native dialect is a northern dialect, higher agreement is expected between her and the northern judges. However, Mandarin adult listeners all showed better than chance agreement. No significant main effect of listener dialect background is found (Figure 2.3: 86% for North, 78% for South).

Figure 2.3 Agreement on tone patterns in infants' babble vocalizations by listener dialect background



These findings also provide evidence that tones can be accurately and reliably categorized by ear as early as infants' babbling period. These results indirectly support Wong (2012), who found that judges were able to categorize the tones in 3-year-old children's vocalizations and established that the categorization of tones is a sensitive and reliable method.

2.6. Experiment 2: Acoustic test

In these children's babble vocalizations some tones were identified as those used in adult Mandarin and others were judged as not fitting any Mandarin tone category ('unidentifiable'). Besides the perceptual test we carried out, additional work was needed to further clarify the results, especially with respect to the unidentifiable patterns. The question was how to measure agreement with regards to those tones, which were difficult to identify. Accordingly, here we use acoustic information to assist in validating the identification of tones. Fundamental frequency is the primary acoustic parameter used to characterize the four Mandarin lexical tones. Unlike adult speech, no automatic instrumental analysis has been developed to quantify the pitch trajectory on a syllable in natural infant babble vocalizations. A combination of pitch analysis and human judgement is applied here. The pitch of each target vocalization is selected in PRAAT to generate a pitch trajectory. Each picture of a pitch trajectory is inspected by the author and another phonetician; pictures in which the automatic pitch tracker clearly made a mistake in identifying the F0 are manually corrected.

Experimental participants are required to visually inspect the pitch contour and compare it to an example of the pitch contour of each standard Mandarin lexical tone.

2.6.1 Stimuli

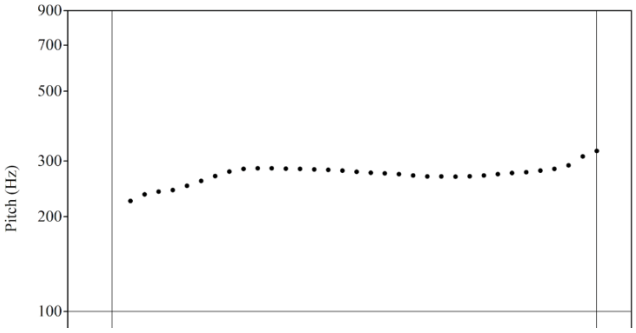
The testing items consisted of 164 pictures showing pitch trajectories in black dots against a white background, which represented 164 utterances: Eighty-two monosyllabic utterances with unidentifiable tones from infants' 9- to 12-month sessions, were extracted. A matching number of monosyllabic vocalizations with identifiable tones were chosen from the infant and session from which the unidentifiable patterns derive were included for comparison. The numbers of unidentifiable tones are not equally represented across the two groups: all seven English infants produced unidentifiable tones in most sessions, but only four Mandarin infants produced such tones in a few sessions (Table 2.6).

Table 2.6 Numbers of unidentifiable tone patterns in monosyllable of each child at each age (%)

	9months	10months	11months	12months
English				
AC	3 (5%)	5 (16%)	4 (16%)	5 (16%)
AH	2 (11%)	0	0	3 (7%)
CL	1 (5%)	3 (15%)	3 (20%)	2 (6%)
CS	2 (10%)	1 (7%)	4 (17%)	2 (12%)
GR	2 (6%)	1 (14%)	0	5 (16%)
JY	1 (8%)	1 (8%)	4 (20%)	6 (15%)
RT	1 (2%)	0	2 (14%)	3 (7%)
Mandarin				
ZZ	2 (4%)	3 (5%)	1 (3%)	1 (3%)
YF	3 (8%)	2 (8%)	0	0
DD	0	1 (4%)	1 (3%)	0

This investigation made use of a picture matching task. Thirty students from the University of York with normal vision took part in the experiment. Each participant was tested singly in a quiet room. Before testing began each participant was screened to ensure that they understood the general testing requirements. All the participants were aware that Mandarin is a tonal language, but none was native speaker and had studied Mandarin or any other tonal languages. Pictures of the pitch trajectory of the four standard Mandarin lexical tones and a fifth option in the form of a question mark were provided as possible choices. All participants were presented with the target pictures at the centre of the screen (see Figure 2.4 for an example) and asked to choose the matched picture from five options: picture of Tone 1, picture of Tone 2, picture of Tone 3, picture of Tone 4 or picture of question mark. Participants were informed that they were being timed, but were not allowed to move on to the next vocalization without making a choice. They were also told it would not be unexpected if many 'unknown' options were chosen.

Figure 2.4 Example of a picture of pitch trajectory on the screen



2.6.2 Statistical approaches

The first analysis was run by participant. For each participant, we calculated an agreement score on all four Mandarin tones and an agreement score for unidentifiable tones. The agreement score was 1 if the participant and I agreed on the tone; was 0 if we didn't agree on the tone. For each participant, we compare the proportion of items which the participants categorized as belonging to the same tone as the researcher, separately for the identifiable and for the unidentifiable tones. After having two percentages of agreement for each participant, an independent T-test was conducted. The independent variable was tone status (2 levels: identifiable [i.e., one of four Mandarin tones], or and unidentifiable). The dependent variable was the proportion of agreement between the participant and the researcher.

The second analysis was run by item. First, for each of the 82 unidentifiable tones and 82 identifiable tones, we calculated the degree of variance in the judgement. That is how variable the 30 participants judge these items. Counting the number of different responses participants gave will tell us how many different tones they assigned each sound file to. Then another independent t-test was conducted, comparing the identifiable and unidentifiable tones on the number of different tones assigned.

2.6.3 Results

The average percentage of agreement was 96% for the identifiable tones, and 58% for the unidentifiable tones (Table 2.7). Less agreement was found in the unidentifiable tones than in the other four Mandarin tones.

Table 2.7 Agreement between participants and researcher in acoustic test

Group Mean

	category	N	Mean	Std. Deviation	Std. Error Mean
Percentage of agreement	identifiable	82	95.98%	11.021%	1.217%
	unidentifiable	82	58.54%	25.038%	2.765%

The independent variable was tone status (2 levels: identifiable [i.e., one of four Mandarin tones], or and unidentifiable). The dependent variable was the proportion of agreement between the participant and the researcher. An independent T-test shows that the difference between conditions was statistically significant (Table 2.8: t= 12.393, df= 162, p< .001, two-tailed).

Table 2.8 Independent T-test of the agreement between two conditions

Independent T-test_Agreement participant vs. researcher

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Percentage of agreement	Equal variances assumed	78.826	.000	12.393	162	.000	37.439%	3.021%	31.473%	43.405%
	Equal variances not assumed			12.393	111.252	.000	37.439%	3.021%	31.453%	43.425%

Regarding the number of different responses to each item given by the different participants, there are only one to two different responses for most identifiable tones, but three to four different responses for many of the unidentifiable tones. This means that for the unidentifiable tones not only did the participants differ in their judgment from the researcher, but they also tended to differ more from each other than for the identifiable ones, as shown in the significant difference identified in the independent T-test (t= -17.703, df= 162, p< .001, two-tailed).

Table 2.9 Independent T-test comparing the number of different responses among participants between two conditions

Independent T-test_responses among participants

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Num_diff_responses	Equal variances assumed	.241	.624	-17.703	162	.000	-1.744	.099	-1.938	-1.549
	Equal variances not assumed			-17.703	146.534	.000	-1.744	.099	-1.939	-1.549

The result indicates the tones that the children produced could be clearly justified as either identifiable or unidentifiable, based on the judgements of the participants. For the identifiable tones agreement on the tone between participants and the researcher was very high. For the unidentifiable tones agreement was low and uncertainty was higher among the judges.

2.7 Analyses of Mandarin adult input speech

To provide an adequate sample of the speech the infants are exposed to, the input of two groups of Mandarin-speaking mothers addressing their children were analysed. We transcribed the mothers' speech to their infants in recordings made when the infants were 12 months old. We calculated the incidence of tone within monosyllabic and disyllabic content words in running speech. All nouns, verbs, adjectives, adverbs and onomatopoeia were treated as content words. Since the incidence of certain tones might be biased by particular lexical items, we looked at type frequency; that is, we count the tone pattern of each lexical item only once.

2.7.1 Monosyllables in Mandarin input speech

As we can see from Table 2.10, although Tone 4 is the most often used tone in monosyllables in the mothers' speech overall, it does not dominate to any great extent; all four tones are produced to a similar extent (above 20%). The standard deviation shows that the distributions of each 'monosyllabic tone' is similar across the mothers' input speech, which means no large individual differences were detected among the mothers.

Table 2.10 Use of tone types in Mandarin-speaking mothers' monosyllabic words in a single 12-month session (NCM= North China Mandarin-speaking mothers, YM=York Mandarin-speaking mothers)

	Tone 1	**Tone 2**	**Tone 3**	**Tone 4**
NCM				
DD	25%	23%	14%	38%
JD	22%	30%	19%	30%
TZ	25%	19%	22%	33%
WZ	28%	22%	25%	26%
SL	24%	17%	24%	34%
KL	24%	21%	28%	28%
LL	31%	15%	29%	25%
WW	25%	18%	29%	29%
QQ	22%	19%	28%	31%
Mean	25%	20%	24%	30%
Sd	3%	4%	5%	4%
YM				
Zuozuo	22%	22%	30%	26%
Weilun	27%	19%	35%	19%
Keke	23%	27%	18%	32%

Yifan	33%	24%	18%	24%
Didi	23%	27%	20%	30%
Xinyu	27%	13%	29%	31%
Yiyi	22%	30%	26%	22%
Shi	25%	18%	23%	34%
Mean	25%	23%	25%	27%
Sd	4%	6%	6%	5%

4.7.2 Disyllables in Mandarin input speech

Disyllables are more variable, so we selected the use of tone patterns used more than 10% and looked to see how many caregivers had the same preference.

Table 2.11 Number of mothers who produced a given tone pattern in more than 10% of their disyllabic words in a single 12-month session (in the order of most use in NCM group)

Groups	T4-T0	T2-T0	T2-T4	T4-T4	T1-T0	T3-T0	T4-T2
NCM	6	6	4	4	3	3	2
YM	7	5	2	3	4	2	3

Despite the differences between the two Mandarin maternal groups, the most frequent tonal patterns used by mothers in disyllables were Tone4-Tone0 (a falling tone followed by a neutral tone) and Tone2-Tone0 (a rising tone followed by a neutral tone). The other five patterns were preferred by at least some mothers. In summary, the preferred patterns were a full tone (either Tone1, Tone2, Tone3 or Tone4) followed by a neutral tone and any Tone2 and Tone4 combination. There are 19 possible tone combinations in dissyllables in Mandarin. Tones which are not illustrated in the table were rarely seen in these samples of input speech.

2.7.3 Syllable structures in Mandarin input speech

The profiles of word length in syllables in input speech are presented. We can see from Table 2.12 that disyllables dominate the speech of all the Mandarin-speaking mothers.

Table 2.12 Distribution of monosyllables and disyllables in mothers' speech in a single 12-month session

NCM	Monosyllables	Disyllables	YM	Monosyllables	Disyllables
DD	35%(44/124)	65%(80/124)	Zuozuo	43%(39/91)	57%(52/91)
JD	36%(24/67)	64%(43/67)	Yifan	41%(46/111)	59%(65/111)
TZ	47%(36/77)	53%(41/77)	Keke	43%(51/114)	57%(63/114)
WZ	40%(65/161)	60%(96/161)	Xinyu	52%(49/94)	48%(45/94)
KL	33%(23/69)	67%(46/69)	Didi	48%(39/81)	52%(42/81)
SL	32%(29/90)	68%(61/90)	Yiyi	49%(51/104)	51%(53/104)
LL	31%(65/209)	69%(144/209)	Shi	48%(64/133)	52%(69/133)
WW	35%(84/241)	65%(157/241)	PingAn	46%(49/107)	54%(58/107)
QQ	41%(34/82)	59%(48/82)			
Mean	37%	63%		46%	54%
Sd	5%	5%		4%	4%

The vocalization lengths that the children produced do not match typical adult forms produced in running speech. In Mandarin spoken language, words tend to be disyllabic (around 60%). Children's early vocalizations are largely restricted to one-syllable forms.

2.7.4 York Mandarin mother-child dyads

We have seen there are individual differences in terms of tone use in both groups of Mandarin-learning children. It is of interest to ascertain where these individual differences come from. Therefore, we are concerned with the differential effect on Mandarin children of hearing different samples of input speech in the same Mandarin community. That is, we want to know if the tone production of the individual infant is influenced by his or her own mother's input speech. For example, does a child who produces a high proportion of Tone 2 have a mother who stands out from the others for her high use of Tone 2 in input speech?

To test the similarity between direct maternal and child tone patterns, we applied the distance matrix approach, which is a distance-function permutation test adapted from Vihman et al. (1994). In this kind of test, smaller differences between each child and his or her own mother than the differences between each child and the other mothers can be taken as support for individual differences originating from different inputs. Such a finding would mean the children are learning specifically from their own their mothers rather than from what they hear more generally, or in other words, that individual differences between the infants are caused by individual differences between the mothers. In our test, the York Mandarin infants and their mothers' tone production at 12 months are sampled. We computed the distance matrix between the proportions of mother's use of each of the four lexical tones and the proportions of each of the four tones in the child's vocalizations. We first characterize each child and each mother by the proportion of his or her use of each tone. We compute the difference between each child and each mother in their proportion of use of each tone (e.g., For Child 1 and Mother 1: Tone 1 in Child 1 minus Mother 1, Tone 2 in Child 1 minus Mother 1, Tone 3 in Child 1 minus Mother 1, Tone 4 in Child 1 minus Mother 1. The same distances are then calculated for Child 1 and Mother 2, Child 1 and Mother 3 and so on for all children and all mothers). Then we square the corresponding differences, to create Table 2.13.

Table 2.13 Difference between each child and mother for four Mandarin tones

	Difference_ Tone1	Difference_ Tone2	Difference_ Tone3	Difference_ Tone4
Child1Mother1	0.00	0.00	0.09	0.04
Child1Mother2	0.00	0.01	0.05	0.01
Child1Mother3	0.00	0.01	0.12	0.07
Child1Mother4	0.01	0.00	0.03	0.04
Child1Mother5	0.00	0.00	0.07	0.05
Child1Mother6	0.05	0.00	0.03	0.02
Child1Mother7	0.00	0.00	0.04	0.02
Child1Mother8	0.00	0.02	0.08	0.02
Child2Mother1	0.03	0.03	0.09	0.10
Child2Mother2	0.02	0.02	0.05	0.06
Child2Mother3	0.01	0.02	0.12	0.15
Child2Mother4	0.00	0.04	0.03	0.11
Child2Mother5	0.03	0.07	0.07	0.13
Child2Mother6	0.02	0.05	0.03	0.07
Child2Mother7	0.02	0.05	0.04	0.08

Child2Mother8	0.01	0.01	0.08	0.07
Child3Mother1	0.03	0.00	0.09	0.02
Child3Mother2	0.02	0.00	0.05	0.00
Child3Mother3	0.01	0.00	0.12	0.04
Child3Mother4	0.00	0.00	0.03	0.02
Child3Mother5	0.03	0.01	0.07	0.03
Child3Mother6	0.03	0.00	0.03	0.01
Child3Mother7	0.02	0.00	0.04	0.01
Child3Mother8	0.02	0.01	0.08	0.01
Child4Mother1	0.00	0.01	0.09	0.18
Child4Mother2	0.00	0.01	0.05	0.12
Child4Mother3	0.00	0.01	0.12	0.25
Child4Mother4	0.02	0.02	0.03	0.20
Child4Mother5	0.00	0.04	0.07	0.22
Child4Mother6	0.00	0.03	0.03	0.13
Child4Mother7	0.00	0.03	0.04	0.15
Child4Mother8	0.00	0.00	0.08	0.14
Child5Mother1	0.00	0.00	0.09	0.13
Child5Mother2	0.01	0.00	0.05	0.08
Child5Mother3	0.01	0.00	0.12	0.19
Child5Mother4	0.03	0.00	0.03	0.14
Child5Mother5	0.00	0.01	0.07	0.16
Child5Mother6	0.00	0.00	0.03	0.09
Child5Mother7	0.00	0.00	0.04	0.10
Child5Mother8	0.01	0.01	0.08	0.10
Child6Mother1	0.01	0.00	0.09	0.06
Child6Mother2	0.00	0.00	0.05	0.03
Child6Mother3	0.00	0.00	0.12	0.11
Child6Mother4	0.00	0.00	0.03	0.07
Child6Mother5	0.01	0.02	0.07	0.09
Child6Mother6	0.01	0.01	0.03	0.04
Child6Mother7	0.00	0.01	0.04	0.05
Child6Mother8	0.00	0.00	0.08	0.04
Child7Mother1	0.03	0.05	0.09	0.12
Child7Mother2	0.02	0.03	0.05	0.07
Child7Mother3	0.02	0.04	0.12	0.17
Child7Mother4	0.00	0.06	0.03	0.13
Child7Mother5	0.03	0.09	0.07	0.15
Child7Mother6	0.03	0.07	0.03	0.08
Child7Mother7	0.03	0.07	0.04	0.09
Child7Mother8	0.02	0.02	0.08	0.00
Child8Mother1	0.02	0.02	0.09	0.08
Child8Mother2	0.02	0.01	0.05	0.04

Child8Mother3	0.01	0.01	0.12	0.12
Child8Mother4	0.00	0.03	0.03	0.09
Child8Mother5	0.02	0.05	0.07	0.11
Child8Mother6	0.02	0.04	0.03	0.05
Child8Mother7	0.02	0.03	0.04	0.06
Child8Mother8	0.01	0.00	0.08	0.05

Then we sum the four squared distances of the four Mandarin tones for each Child-Mother pair and take the square root of this sum. The resulting number expresses the distance between Child X and Mother Y. We thus create an 8*8 matrix of distances for all Child-Mother pairs in terms of the relative use of four Mandarin tones (see Table 2.14). We then transform this table into a table of ranks, where distances are ranked within each child's row from 1 (smallest) to 8 (highest) (see Table 2.15).

Table 2.14. Matrix of distances for all Child-Mother pairs in four Mandarin tones

	Mother 1	Mother 2	Mother 3	Mother 4	Mother 5	Mother 6	Mother 7	Mother 8
Child 1	0.36	0.27	0.45	0.29	0.35	0.23	0.25	0.35
Child 2	0.50	0.38	0.56	0.43	0.54	0.42	0.44	0.42
Child 3	0.37	0.27	0.27	0.24	0.37	0.26	0.28	0.34
Child 4	0.53	0.42	0.42	0.51	0.57	0.44	0.46	0.48
Child 5	0.47	0.37	0.37	0.45	0.49	0.36	0.39	0.44
Child 6	0.41	0.29	0.29	0.33	0.42	0.29	0.31	0.36
Child 7	0.54	0.42	0.42	0.48	0.59	0.47	0.48	0.45
Child 8	0.46	0.34	0.34	0.39	0.50	0.38	0.39	0.39

Table 2.15 Rank of the matrix of distances

	Mother 1	Mother 2	Mother 3	Mother 4	Mother 5	Mother 6	Mother 7	Mother 8	
Child 1	7.0	3.0	8.0	4.0	6.0	1.0	2.0	5.0	4.5
Child 2	6.0	1.0	8.0	4.0	7.0	3.0	5.0	2.0	4.5
Child 3	8.0	3.5	3.5	1.0	7.0	2.0	5.0	6.0	4.5
Child 4	7.0	1.5	1.5	6.0	8.0	3.0	4.0	5.0	4.5
Child 5	7.0	2.5	2.5	6.0	8.0	1.0	4.0	5.0	4.5
Child 6	7.0	1.5	1.5	5.0	8.0	3.0	4.0	6.0	4.5
Child 7	7.0	1.5	1.5	5.0	8.0	4.0	6.0	3.0	4.5
Child 8	7.0	1.5	1.5	4.0	8.0	3.0	6.0	5.0	4.5

If the child's tone production is most influenced by his or her own mother's use, then the distance between each child and their own mother will be the smallest in the row, and the diagonal entries will be the smallest in the matrix. On the other hand, if the mother's use of tone categories had no effect on her child's use, then diagonal matrices will rank 4.5 on average (the average rank in a scale from 1 to 8). The actual mean ranks for the diagonal cells were: 7.0, 1.0, 3.5, 6.0, 8.0, 3.0, 6.0, 5.0. A single sample T test was conducted to determine whether these means were significantly smaller than 4.5. The difference was not significant ($p > .05$).

2.8 Discussion and Conclusion

This chapter sought to assess the presence of ambient language influence in babbling. We followed a number of steps to arrive at empirical results. Mandarin- and English-learning infants' babble vocalizations were compared cross-linguistically regarding the production of tones. Mandarin is a tone language and has a systematic distinction among tones at the lexical level, whereas English does not distinguish different tones. First, we drew on transcription data from children acquiring each of the languages. Previous studies reported ambient language effects in babbling cross-sectionally; here we attempted to arrive at a more reliable estimate. We followed infants longitudinally from 9 month to 12 months in both languages and made use of frequency and stability criteria (McCune & Vihman, 2001),

The transcription was then tested in two reliability experiments. Forty adult Mandarin speakers listened to randomly selected infant vocalizations (n=140) to determine the tones. There was good agreement for utterances from each language background (Mandarin: 85%; English: 81%), which means tones produced in babbling are categorizable. The complex tonal patterns which could not be included in the statistical analysis were categorized by thirty non-Mandarin-speaking judges. The judges showed overall reliable identification, which supported our impressionistic auditory transcription.

Both similarities and specific language characteristics have been found here to underlie the babbling of Mandarin- and English-acquiring infants. It has been suggested that biological constraints give children's output considerable similarity across all speech communities – and indeed a high percentage of falling tone (Tone 4 in Mandarin) is found in many language groups (Locke, 1983). This biological model is confirmed in our data in that both Mandarin- and English-acquiring infants produced frequent Tone 4. 'Biological' here means constraints on phonation, control of the rate of vocal cord vibration. Changes in pitch are made by manipulating tension in the vocal folds. The larynx as a valve manages the flow of air into and out of the lungs, and consists of three different types of cartilage: the thyroid, the cricoid and a pair of arytenoid, which work together to enable the vocal folds to vibrate in different ways: tensing to produce high pitch (e.g. Tone 1 in Mandarin) or relaxing to result in falling pitch (e.g. Tone 1; Yip, 2002). Pressure goes down naturally as one speaks, which slows down the rate of vocal fold vibration, which eventually results in a falling pitch. Therefore, releasing the vocal folds naturally to produce Tone 4 is easy for children to learn to control.

Ambient language effects in babbling were identified through cross-linguistic comparisons of certain aspects. Mandarin-learning infants produced a considerably larger number of Tone 1 syllables while English-learning infants showed different tendency, as their babbling was characterized by a predominance of falling (equal to Tone 4 in Mandarin). This is consistent with previous observations on intonation, to the effect that falling contours (either rise–fall or fall) are the most frequently occurring intonation patterns of English-learning infants in longitudinal studies (Kent & Bauer, 1985; Robb et al., 1989). The high use of Tone 1(level) in Mandarin suggests that infant production already reflects this input tone pattern in the pre-linguistic period. (Note that Tone 1 is essentially non-occurrent in English). Furthermore, the greater use of Tone 1 in the infants living in a monolingual community in North China (as compared…) further suggests that ambient language effect. Meanwhile, the production of

Tone 4 in Mandarin infants matches the percentage of Tone 4 in adult input speech, so that this tendency can also be seen as reflecting the properties of the ambient language environment.

In children's disyllabic vocalizations, recall that the English infants primarily produced Tone 2 + Tone 4, Mandarin infants produced that pattern the most frequently as well, but in addition they produced Tone 1 + Tone 1 and Tone 1 + Tone 4. This tends to reflect the children's language environment, in that English infants produce no Tone 1 in either syllable of their disyllabic words.

The infants exposed to Mandarin were found to produce more categorical tones than the infants exposed to English. Two different tone contours occurring in one monosyllable of infants' vocalizations were also detected perceptually. English infants have more of this complex pattern. This provides subtle evidence of an ambient-language effect: Mandarin infants are exposed to constrained input speech involving no more than one lexical tone on each syllable, while English infants are not. As Mandarin is a tone language, tones characterize every single syllable; therefore, children may be learning to attend to pitch changes at the syllable level. On the other hand, for a non-tone language pitch correlates only with intonation over longer units; thus children do not need to learn to attend to or maintain a single pitch value at the syllable level.

This study is also the first to directly investigate Mandarin-learning infants' tone use in babbling. Although the between-group differences are statistically significant, Mandarin infants' tone use was also found to differ by child. For example, some Mandarin-learning infants produced less Tone 1 and more Tone 2. We analysed potential sources for this variability: differences in the mothers' input speech and in individual mother-child dyads. If parental speech constitutes the major source of individual differences, there should be noticeable variation in Mandarin mothers' speech and high similarity between each child and his or her own mother. However, our results failed to support that proposal. Four tones were produced to a relatively equal extent overall across all mothers. Children's differences in their tone production choices are not related to the specific input speech to which they are exposed. Therefore, in our view, the ambient language shapes the vocalizations of the group of children as a whole, while each child filters the speech on the basis of his or her own articulatory ability, motor skill or past experience. In the proportion of mono- and disyllables,

we can see that monosyllables characterized most Mandarin-acquiring children's babble vocalizations for all four months investigated here.

3. PROSODIC STRUCTURE AND SEGMENTAL CHALLENGES IN EARLY WORD PRODUCTION

3.1 Introduction

In their second year children usually begin to produce words. Within the word learning period, children share common phonological characteristics: they produce a small number of relatively accurate words at first, and then shift to expanding familiar patterns into more difficult words (Menn & Vihman, 2011). They have difficulty retaining and planning difficult words due to their limited consonant inventory and phonological memory. To cope with the challenge, children learning different ambient languages find different solutions. However, learners acquiring non-European languages have received much less attention. This chapter reports on an observational longitudinal study of Mandarin-learning children at two word-points (around 11 to 17 months of age). It is intended to provide a clear picture of Mandarin children's word production in terms of prosodic structures, consonantal resources and possible templates, to evaluate the advances, challenges and changes from the first words to well into the single-word period. The development of lexical tone as an element of Mandarin phonology is reported separately in the next chapter.

3.1.1 Prosodic structure

Prosodic structure as defined by Selkirk (1980, 1984) refers to a multi-layered structure that expresses rhythmic chunks consisting of different lengths. Syllables are viewed as independent prosodic units. (Waterson, 1971; Menn, 1971; Stampe, 1979; Vihman, 1978; Goldsmith, 1990). Of all the constituents in the prosodic hierarchy, the syllable is the only unit that has been investigated extensively in child language research. The role of the syllable in early phonological acquisition has been documented in the literature. In perception studies infants have been found to be sensitive to syllable boundaries (Jusczyk, 1997; Eimas & Miller, 1980) and early production studies have identified a prevalence of CV forms and other phonotactic constraints (Menn, 1971; Smith, 1973; Vihman, 1992,1996; Bernhardt & Stemberger, 1998).

Syllable types vary across languages. Some languages permit only CV syllable structure, whereas others allow more complex syllable structures with onset and coda clusters, for

example. Mandarin syllable structure is relatively simple compared with English. Syllable and word shapes include V (or VV/VVV; VV = diphthong; VVV = triphthong), CV (or CVV/CVVV), CVC (or VC/CVVC), CVCV (or CVVCVV) and CVCCVC (e.g. [koŋ3loŋ2] 'dinosaur'). Only nasals can serve as syllable-final consonants.

The prosodic structures that this chapter will focus on are those produced in the early word period by Mandarin-learning children. We will present the main types of prosodic patterns found, then consider likely sources for those patterns, including constraints on vocal production, aspects of input speech, or the interaction of the two. For example, we will consider whether the structures children use reflect the most frequent patterns in the input or a universal pattern like open CV syllables.

3.1.2 Segmental challenge

There are cross-linguistic similarities in children's first identifiable words both segmentally and in terms of prosodic structure. Children mostly produce early words accurately with only one or two syllables in length, with open syllables and a single place or manner of articulation of consonants within the word, and target-limited vowels. Menn & Vihman (2011) provide evidence from 49 children learning 10 languages.

This characterisation of first words suggests constraints due to children's early articulatory insufficiency. Children still control tongue posture at the whole-utterance level (frame-dominance hypothesis, MacNeilage & Davis 1990): The child's tongue is large in relation to the size of the vocal tract and occupies a more anterior position in the oral cavity. The position of the tongue relative to the jaw remains relatively constant, and tongue and jaw move together as a single unit. These simple ballistic movements result in a small consonantal repertoire. Children's first words also depend on their communicative needs; children select for production some routine or often-used sound patterns which fit their own vocal patterns.

Phonological memory, which develops through the emergent use of word forms in production, is another challenge for children producing early words that resemble their targets (Keren-Portnoy et al., 2010). Words have a memory representation which can be conceptualised as a cluster of exemplars (Bybee 2001, Pierrehumbert 2001). Producing a

single gesture repeatedly is less challenging than planning and sequencing multiple distinct gestures (Pater 1997). Children have difficulty combining multiple discrete gestures into a complex sequence, consonant harmony as result is major place of articulation. Children in Keren-Portnoy et al. (2010) repeated familiar words more successfully than nonwords, even when the real words included less familiar consonants.

Given limited articulatory and representational skill, producing multiple gestures would be much harder than repeatedly producing a single gesture. Producing words with final consonants or more than one consonant type are challenges European language-learning children face. To manage their limited capacities, children find different strategies; for instance, consonant harmony and reduplication are two well-known preferences (e.g., Vihman, 1978; Stemberger & Stoel-Gammon, 1994).

However, no studies have examined child-specific phonological patterns for children exposed to Mandarin. Studies of phonological development in Mandarin have all investigated the order and age of acquisition of phonemes and segmental or tonal errors (e.g. Zhu, 2002). Therefore, we do not know what governs children's choice of a pattern that they substitute for the adult pattern, or what challenges Mandarin children may face. This analysis will address these questions.

3.1.3 The current study

One goal of this chapter is to examine the challenges of prosodic structure and segmental pattern that Mandarin children face in the process of learning to produce words. Specifically, this chapter will identify the prosodic structures and consonant inventory of Mandarin-learning children, both in their 4-word point and 25-word point. We also consider whether it is possible to identify templates in Mandarin-learning children by the end of the single-word period (25-word point). The second goal is to see, from the point of view of each child's own independent system, to what extent universal constraints affect their word production patterns and to what extent the ambient language shapes those patterns. To achieve this, we will compare these Mandarin-speaking children to children learning different languages (drawing on Vihman, 2019) to see similarities or differences. We will then investigate the input language to see how the adult language shapes children's early words.

3.2 Method

The present chapter is a follow-up to the babble study of the children from Mandarin-speaking families in York in Chapter 2. Eight children (six females, two male) were followed longitudinally from age 1;0 to 1;5. The same recording methods were applied; children were engaged in largely unstructured play at home with their mothers. Two word-points were used for sampling the data in a way comparable to data from other languages. Three children did not meet the 25-word point by the end of data collection (One went back to China at about age 1;2 and two left at about age 1;4 due to their fathers' job relocation); therefore, only five children's data sets are available for analysis at the 25-word point.

Words were identified using the criteria proposed by Vihman and McCune (1994) and criterion was given for each child's vocalizations. A word is not required to be pronounced exactly like the adult form. A word differing from adult target could still be acceptable if it is used in the appropriate contexts and in a consistent way. Common phonological reduction rules were considered (see Vihman et al. 1985). Target words were defined as the adult pronunciation of the particular word which the child was attempting to produce.

The Mandarin CDI was administered in this study on a monthly basis to provide a parental estimate of vocabulary size (Hao, Shu, Xing and Li, 2008: the Infant part is from 12 to 16 months and the Toddler part from 17 to 30 months). Table 3.1 shows the word use reported by parents at each child's lexical points.

Table 3.1 Maternal reports of Mandarin children's words use at two lexical points

	4-word point	25-word point
Keke	16	92
Xinyu	11	88
Yifan	5	N/A
Yiyi	13	63
Zuozuo	8	N/A

Didi	9	57
Shi	7	30
Weilun	8	N/A

3.3 Data analysis and results of children's production

3.3.1 Prosodic structure in Mandarin-acquiring children at 4-word point

The chapter will start by discussing preferred prosodic structures in the first recorded words (4wp) of eight Mandarin-learning children. The data give some idea of the range of Mandarin children's ages at first words (from 11 to 13 months), preferred first word forms and their onset of phonological development. By comparing prosodic structures across children and looking at individual differences, this analysis will reflect children's uses of the most frequent adult-like patterns, their competence in mapping input word forms onto their existing vocal patterns or the reverse and their favoured structures. Table 3.2 presents the first recorded words identified, with all of their variant forms.

Table 3.2 Children's words at 4wp. The data sets are ordered by child age; textual references to individual children retain that ordering.

Child name (age)	adult form	gloss	child form
1. Keke (11mos.)			
	/ta1/	build	[ta1], [ta4]
	/tiau4/	drop	[tou1], [tiu4]
	/kei3/	give	[ke4](2), [kɛ2ke4]
	/pa4/, /pa4pa0/	daddy	[pa4], [pa1pa0], [pa4pa0], [pa4pa4]
	/nai1nai0/	(breast) milk	[na1na0]
	/ka1ka1/	quack	[ka1ka4](3), [kɤ1ka4]
	/tau2tau3/	fall down	[ta1tu4], [tau4tau1], [tu4tu4]
2. Xinyu (11mos.)			
	/wei2/	hello (on the phone)	[wei2], [wei4](2), [wa4](2), [wa1wa0]
	/wo3/	me	[wo4](2), [wo4wo4]
	/waŋ1/	woof	[wa1]

/miau1/	meow	[mɛ1](2), [mi4], [miɛ1me4], [miɛ1mi4]
/pai1pai1/	byebye	[pa1pa0](2), [pɛ1pɛ4](2), [pa4pa4](2)

3. Yifan (12mos.)

/ta4/	big	[ta4]
/pai1pai1/	byebye	[pa1](4), [pa1 pʷa 1], [pai1pai1]
/tɕiɛ3tɕiɛ0/	sister	[tɕi1ɕiɛ4](2), [tɕiɛ1kɛ4]
/kei2ni0/	there you go	[ke1jiɛ4], [ka4jiɛ4]

4. Yiyi (12mos.)

/ta3/	hit	[ta1], [ta4](2), [ta1ta4](2)
/pau4/	hug	[pa1](5), [pa4](5)
/ma1/, /ma1ma0/	mummy	[ma1], [ma1ma4]
/pai1pai1/	byebye	[pa1pa1] (3)
/ɕiɛ4ɕiɛ0/	thanks	[ɕiɛ4ɕiɛ0], [ɕiɛ1ɕiɛ4], [ɕia1ɕia1], [ɕia1ɕia4] [ɕia2ɕia4](2), [ɕiɛ1ɕia1], [ɕiɛ4ɕi1]

5. Zuozuo (13mos.)

/xua1/	flower	[wa1], [pʷa1]
/bu4/;/bu2jau4/	no	[pu4] ; [pu1], [pu2](2)
/mau2/	peekaboo	[ma2], [mau2]
/wo3/	me	[wa4wo4] [wo1]
/pa4/, /pa4pa0/	daddy	[pa4], [pa2pa2]
/ma1ma0/	mummy	[ma1ma0](8), [ma1ma4], [ma1ma2]

6. Didi (13mos.)

/kei3/; /kei2/([ni0])	give	[kʸe2], [kɛ1], [kɛ2], [ka1], [ka2]
/tu1tu1/	beep	[tu1tu1], [tʲu4tʲu1]
/ma1ma0/	mummy	[ma1ma0](6)
/pa4pa0/	daddy	[pa1pa0], [pa4pa0](2), [pa4pa4]
/tɕiɛ3tɕiɛ0/	sister	[tiɛ1tɛ4], [tɕiɛ2](2)
/na4na0/	sister's name	[na4na0](6)
/ɕiɛ4ɕiɛ0/	thanks	[tə4tə0], [tɕʰiɛ4tɛ1], [tɕə4tɕɛ4], [tɕɛ4tɕiu4]

7. Shi (13mos.)

/wu1/	choochoo	[ʯu1](4)

/ta4/	big	[ta4], [tʸɛ4], [tɛ4]
/na2/	take	[na1na1]
/ma1ma0/	mummy	[ma1ma0](2), [ma2ma4]
/pa4pa0/	daddy	[pa4pa0](2)
/pa1pa1/	woof	[pa1pa1]

8. Weilun (13mos.)

/la1/	pull	[la4] (2)
/ti1ti1/	beep	[tɛ1], [ti1ta4], [ta1ta4], [tɛ4ta1], [ta4ta1]
/ma1ma0/	mummy	[ma1ma0](2), [ma1ma4]
/mou1mou1/	moo	[ma4mu4], [mu4ma4]
/ta3/	open	[ta1ta0], [ta1ta4]

We see from Table 3.2 that of the 32 target words, eight are onomatopoeia or non-verbal sounds (quack-quack, meow, moo, woof [*2], beep [*2], choochoo, peekaboo [*2]); most are terms fpr family members (e.g. mummy [*5], daddy [*4], sister [*2] or myself). The target words include no more than two syllables and vary between mono- and disyllables. Almost all forms feature open syllables; only one CVC word (/waŋ1/ woof) is targeted, by a single child. Most of the target words include only a single consonant type (30 out of 32, 94%). We consider the prosodic patterns children produced. The children differ in their preferred patterns (Table 3.3): two children use more monosyllables (Xinyu and Zuozuo), the other six mainly use disyllables. They produce monosyllabic patterns for some targets, disyllabic patterns for others, or vary between the two for the same adult target.

The children's patterns all consist of open syllables and mainly have monophthongs. Four children produce CVV one word each. Diphthongs are not well distinguished from monophthongs in disyllabic structure. Only two children produce more than one supraglottal consonant type within a word: Yifan produces [tɕiɛ1kɛ4] among her variants of /tɕiɛ3tɕiɛ0/ 'sister', the velar /k/ perhaps reflecting effortful exploration of the Mandarin affricate /tɕ/; the same account would apply to Didi's [tɕʰiɛ4tɛ1] for /ɕiɛ4ɕiɛ0/ 'thanks'. Meanwhile Yifan also targets another variegated word, /kei2ni0/ 'there you go', but she produces it as [kɛ1jiɛ0], [ka4jiɛ0] with only a single supraglottal consonant.

Table 3.3 Summary of prosodic structures (ordered by proportion of disyllables)

	Monosyllables			Disyllables		
	CV	CVV	CVC	$C1VV_0C1VV_0$	C1VC2V	Single C
Didi	11% (1/9)	11% (1/9)		67% (6/9)	11% (1/9)	
Yifan	29% (2/7)			43% (3/7)	14% (1/7)	14% (1/7)
Shi	33% (2/6)			67% (4/6)		
Weilun	33% (2/6)			67% (4/6)		
Keke	30% (3/10)	10% (1/10)		60% (6/10)		
Yiyi	43% (3/7)			57% (4/7)		
Xinyu	44% (4/9)	11% (1/9)		44% (4/9)		
Zuozuo	50% (4/8)	13% (1/8)		38% (3/8)		
Mean	34%	6%		55%	3%	2%
Sd	12.0%	1.3%		12.0%	2.1%	

3.3.2 Comparison in target words at 4-word point

Now we look at the length in syllables of target words and the proportion of matches across Mandarin children in a single group, and then compare them as a group to the different language groups in Vihman (2019) to get a cross-linguistically comparative perspective on word shapes. Child variants of different lengths are treated as a match if at least one variant matches the target in length.

Table 3.4 Length in target syllables of 7 Mandarin children and proportion of matches (ordered by numbers of disyllables).

Name	Length in target syllables		Total words (matched)	Child form
	one	two		Proportion matching in length
Keke	3	4	7 (7)	1.00
Xinyu	3	2	5 (5)	1.00
Yifan	1	3	4 (4)	1.00
Yiyi	2	3	5 (5)	1.00
Zuozuo	4	2	6 (5)	.83
Didi	1	6	7 (7)	1.00
Shi	3	3	6 (5)	.83
Weilun	2	3	5 (4)	.80
Total	19	26	45 (42)	.93

Although all seven Mandarin children limited their targets to monosyllables and disyllables in Table 3.4, they varied in the proportion of different word shapes. Two children (Xinyu and Zuozuo) attempted more monosyllables than disyllables, two children (Yifan and Didi) attempted considerably more disyllables than monosyllables, while the remaining four children target similar numbers of monosyllables and disyllables. There are both monosyllables and disyllables for the same target words in some occasions, but at least one form matches the target. Therefore, Mandarin children generally have no problem with the syllable length they attempt.

Table 3.5 Length in target syllables of 8 language groups (including Mandarin) and proportion of matches (ordered by proportion of disyllables).

	one	two	three		Proportion matching in length
Italian	3 (0.07)	37 (0.90)	1 (0.02)	41	1.00
Finnish	2 (0.08)	22 (0.85)	0 (0.00)	24	.92
Welsh	6 (0.25)	17 (0.71)	1 (0.04)	24	.96
Estonian	9 (0.35)	17 (0.65)	0 (0.00)	26	.81
Mandarin	19 (0.42)	26 (0.57)	0 (0.00)	45	.93
French	9 (0.45)	11 (0.55)	0 (0.00)	20	.90
US English	14 (0.48)	15 (0.52)	0 (0.00)	29	.83
UK English	22 (0.47)	18 (0.38)	4 (0.09)	47	.88

We compared our Mandarin data to 39 children in six language groups (Table 3.5): British English (8), Estonian (5), Finnish (5), French (4), Italian (7), Welsh (4) and American English (6) in Vihman (2019, and found that children from all language groups attempted mainly mono- or disyllables at the 4-word point. Like children learning six other languages (except UK English), Mandarin-learning children also attempt somewhat more disyllables than monosyllables. The level of match between child forms and target forms is different but high, which means children from all language groups are highly accurately producing the length of the adult forms the 4-word point (Table 3.5).

We can see that the target words are very similar and the productions are generally accurate in terms of syllable structures among Mandarin children. The prosodic structure of Mandarin children's first words and the target words in children learning other languages reveals the cross-linguistic similarity.

3.3.3 Segmental resources in Mandarin-acquiring children at 4-word point

This section presents the main types of segmental patterns in the 4wp production of Mandarin children. Although the segments that the children produce are variable by child, stops and nasals are the most common patterns for all children (see Appendix III). As for place of articulation, all the children produce primarily bilabial or coronal segments. Only three children produce velars in four words but all match the consonant of the targets: Keke produces [k] in two words, both monosyllable and disyllable; Yifan and Didi in just one. In terms of manner of articulation, children mainly produce stops and nasals. Fricatives and affricates occur in the first words of only three children. Two children produce affricate [tɕ], both for the target word /tɕiɛ3tɕiɛ0/ 'sister'. One child produces fricative [ɕ] for the target form /ɕiɛ4ɕiɛ0/ 'thanks'.

The vowels are largely drawn from the front portions of the vowel triangle, as transcribed. Mid/lower front vowels (e.g. [ɛ], [a]) are more frequently used in both monosyllabic and disyllabic words. High front vowel [i] is used in three children's variants of words. High back vowel [u] is also produced by some children. Diphthongs start to emerge in these first words, with the most common one being [iɛ] for particular targets (Yifan and Didi /tɕiɛ3tɕiɛ/ 'sister', Yiyi /ɕiɛ4ɕiɛ0/ 'thanks').

Due to the late acquisition and the challenge of accurate transcription, voicing is disregarded here. Disregarding tones at this point, these Mandarin-learning children's very first words are relatively accurate. Their failures to match the targets are more due to omission than to substitution. Although there are individual differences, Mandarin children's word forms are limited and broadly similar.

Mandarin children's productions are limited in syllable length and segments. These constraints might be due to articulatory ability: The children may not yet have the articulatory control to produce codas, velars and fricatives/affricates, and there are also memory limitations, the lack of lexical experience that prevents the child from successfully deploying memory in the earliest period (Macken, 1995, p. 684).

3.3.4 Prosodic structures in Mandarin-acquiring children at 25wp

Now we turn to the period when the Mandarin children spontaneously produce at least 25

different words in a session, to see the changes over time. We will start by looking the prosodic structures used by each child, and then describe general tendencies and any individual differences. The analyses include the number of words produced either spontaneously or in imitation and the number of different prosodic variants for each target word.

Criterion of prosodic structure, word variants and 'selected' and 'adapted' words at 25wp
The standards for classifying prosodic structures are listed here: Prosodic structures are treated as distinct only if they reach a minimum of 10% of the entire set of classifiable variants within a session (see Vihman, 2016). The children's productions have been sorted into structures by length in syllables, i.e., CV and CVCV. Once the total number of variants in recognisable structures reach the 10% criterion, they are analysed more finely, to identify the proportion of forms with diphthong (VV), coda, or with and without harmony. Where two such structures fail to reach 10%, they are later combined. Word forms differing in voicing or the quality of vowel (or tones) but sharing the same prosodic structure are treated as one variant (e.g., Keke, this /ʈʂɤ4/: [ʈʂɤ4], [kɤ4]). However, the forms of the same target word are counted separately if they fit different prosodic structures (e.g., Keke, /bau4bau0/ 'hug ': CVCV [ba4ba0], CVVCVV [bau4bau0]), which means that a single target word is counted more than once only if the child produces it with variants that match different structures. When new structures are identified with further analysis, the total variants are counted again and the 10% criterion checked again as well.

The procedures for distinguishing variants are as follows. First, all variants produced in a session are identified. Second, variants differing in syllable shape are distinguished to establish the prosodic structures (for example, CV, CVV and CVC for monosyllables, VCV, C_1VC_1V, C_1VVC_1VV, C_1VC_2V and C_1VVC_2VV for disyllables). Third, structures are combined if they add up to fewer than 10%. For each structure, the proportion of all variants will change for different analysis purposes. Here, in the following analysis, disyllabic variegation with monophthong and diphthong are not distinguished as separate structures due to insufficient occurrence, so the variants are combined as $C_1VV_0C_2VV_0$.

Selected and Adapted criteria are applied in each prosodic structure of each child. This means that the target form reflects the child's existing motoric experience and the shapes that the child has mastered so far. Regardless of segments, only prosodic structures are considered,

for example, Keke, CV: [kɤ4] for /ʈʂɤ4/ ‘this’; CVV: [fei4] for /xui4/ ‘can’. Word forms that are missing a structure that the child never produces could also be treated as selected, depending on each child’s existing phonological knowledge in the session; for example, words with a missing coda (as we noted earlier there are only nasal codas in Mandarin) are treated as selected for children like Keke, Xinyu and Shi, who do not produce codas at all, but not for Didi and Yiyi, who do produce them. Take Keke, for example: we consider [kʰɤ4] for /kan4/ ‘look’ to be selected. ‘Adapted’ words here are, for example, diphthongs turned into monophthongs, or vice versa, in a monosyllabic or disyllabic structure if a child has both structures (Keke, CV: [kʰo1] for CVV: / kʰai1/ ‘open’); disyllables turned into monosyllables (Keke, CV: [ma1] for /ma1ma0/ ‘mummy’); variegated syllables turned into reduplication (Keke, C_1VC_1V: [kʰɤ3kɤ1] for /xɤ2tsi0/ ‘box’).

The tables below show five children’s word production and prosodic structures. We begin by presenting Keke, in Table 3.7a, followed by some discussion of what the table shows; the remaining children are presented in Tables 3.7b, c, d, e, f and g.

Table 3.7a Keke’s words at 25wp (18 months), ordered by prosodic structures. ‘Selected’ words are close to the target, suggesting possible sources for the template; ‘adapted’ words show changes to the target; C = consonant, V = vowel, C_0 = optional consonant slot

selected words		adapted words	
target	child form	target	child form
CV 9 (.21)			
哇 wa4	wo4	手 ʂou3	kʰo2
看 kʰan4	kʰɤ4	盖 kai4	kɤ1
这 ʈʂɤ4	ʈʂɤ4, kɤ4	开 kʰai1	kʰo1, kʰo4
		妈妈 ma1ma0	ma1
		鞋带 ɕiɛ2tai4	ta4

		衣服 ji1fu2	fu1

C_0VV 6 (.14)

盖 kai4	kɤɑ1, kɤɑ4, kɤə4	宝贝 pau3pei4	pei4
没 mei2	mei2		
会 xui4	fei4		
手 ʂou3	ʂou2		
好 xau3	au3, au4, au2(4)		

C_1VC_1V 11 (.25)

妈妈 ma1ma0	ma1ma0 (8)	拜拜 pai1pai2	pa1pa2 (3), pa4pa2
爸爸 pa4pa0	pa4pa0, pa1pa0 (6)	抱抱 pau4pau0	pa4pa0 (4), pa2pa0 (2)
可可 $k^hɤ2k^hɤ3$	$k^hɤ2k^hɤ3$(2), $k^hɤ1k^hɤ4$	宝宝 pau3pau0	pa2pa0 (3)
		盖盖 kai4kai0	kɤ1kɤ1(2), kɤ1kɤ0
		nai1nai0	nə1nə0
		外婆 wai4p^{h}o2	wa1wa0
		盒子 xɤ2tsi0	$k^hɤ3$kɤ1
		宝贝 pau3pei4	pɤ3pe4

$C_1VVC_1VV_0$ 7 (.16)

mia1mia1	mia4mia4
拜拜 pai1pai2	pai4pai2
抱抱 pau4pau0	pau4pau0
宝宝 pau3pau0	pau2pau0
掉掉 tiau4tiau0	tau1tau4, tau4tau0
谢谢 ɕiɛ4ɕiɛ0	ɕiɛ1ɕiɛ4
休息 ɕio1ɕi0	ɕio1ɕi4

C_1VV_0 C_2VV_0 5 (.11)

瓜子 kua1tsi3	ku1si0		
不掉 pu2tiau4	pu1tau1	谢谢 ɕiɛ4ɕiɛ0	$k^hɤ1tɕ^hiɛ4$, $t^hi1ɕiɛ4$, $t^hi1ɕiɛ0(2)$, $t^hɛ1ɕiɛ0(2)$
老虎 lau2xu3	lou1fu4	鞋鞋 ɕiɛ2ɕiɛ0	$k^hɤ2ɕiɛ2$, $k^hɤ2ɕi2$

single C & glide 6 (.14) or no supraglottal consonants

芋头 ju4t^{h}ou2	ju2tou2
抱我 pau4wo3	pa4we1
我会 wo3xui4	wu1fui1, en1fei4
哎呀 ai1ja0	ai1ɻa0(3), ai1ɻa4
hiya	ai2ja1, ai4ja1

hello	a1lou1, ai1lou1, a1lou2(2), ai4ou2

Keke produced 36 different words spontaneously and a total of 44 variant forms. As VV (one form) is not sufficiently well represented to be included as a separate structure, the variants are combined as C_0VV. Here again, as there is only one instance of C_1VVC_1V, the variant is combined in the diphthongal disyllabic reduplication-$C_1VVC_1VV_0$. In this case, Keke's prosodic structures are divided between 9 CV forms (21%) and 6 C_0VV forms (14%) for monosyllables and 11 CVCV forms (25%), 7 $CVVCVV_0$ forms (16%), 5 $C_1VV_0C_2VV_0$ forms (11%) and 6 single C (14%) for disyllables. C_1VC_1V is Keke's most favoured structure, accounting for 25% of all forms (11 forms), and mainly contains adapted forms: five out of eight are involve a reduction of a diphthong, the other two are consonant-harmony forms for variegated targets ([wa1wa0] for /wai4pʰo2/ 'grandma'; [kʰɤ3kɤ1] for /xɤ2tsi0/ 'box') and one is disyllabic harmony ([pɤ3pe4] for /pau3pei4/ 'baby'). CV is her second favoured structure, accounting for 21% of all forms, with more adapted forms than selected. Keke also makes high use of diphthongs, in both monosyllables and disyllables and primarily includes selected words.

Table 3.7b Didi's words at 25wp (18 months), ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 9 (.22)			
不 pu4	pu1(2)	桌 ʈʂuo1	ʈʂu1, ʈʂu4
车 ʈʂʰɤ1	tʰɤ4	嘎嘎 ka1ka1	ka1(4)
这 ʈʂɤ4	tɤ1(2)	滴滴 ti1ti1	ti1
吃 ʈʂʰi1	ʈʂʰi1(2)	妈妈 ma1ma0	ma1(2)
		弟弟 ti4ti0	ti4(3), ti2

CVV, CVC 5 (.12)

消失又出现 mau2	mɤɚ1
猫 mau1	mau1(6), mau4(6)
给 kei3	kei2(2)
go	kʰou1
汪 waŋ4	waŋ4

C_1VC_1V 13 (.32)

吃饭声音 a2mu4	ma2mu4	宝宝 pau3pau0	pa3pa1, pau3pa1
妈妈 ma1ma0	ma1ma0	在这 tsai4ʈʂe4	ta1tu4, tu2tu4(2), tu4tu4, tu4tu0
爸爸 pa4pa0	pa4pa0(3)	痛痛 tʰʊŋ4tʰʊŋ0	tu4tu4, tu4tu0(2)
打打 ta2ta0	ta2ta0(2), ta4ta0	走走 tsou2tsou0	tu1tu0
滴滴 ti1ti1	te1te1(2)	姐姐 tɕiɛ3tɕiɛ0	tɛ3tɛ0(2)
弟弟 ti4ti0	ti4ti0, ti1ti1(2)	桌子 ʈʂuo1tsi0	tse1tsi0
车车 ʈʂʰɤ1ʈʂʰɤ1	tɤ1tɤ4(2)		

$C_1V\ V_0C_1VV_0$ 3 (.07)

拜拜 pai1pai2	pai1pai2(3), pai2pai2, pai3pai4(2)
宝贝 pau3pei4	pau3pa4(3), pau3pa1(4)

姐姐 tɕiɛ3tɕiɛ0	tɕiɛ3tɕiɛ0, tɕʰi3tɕiɛ1(2), tɕʰi3tɕʰiɛ1, tʰɛ3tʰiɛ1(2)		
$C_1VV_0C_2VV_0$ 5 (.12)			
这个 ʈʂɤ4kɤ4	tɤ1kɤ4, tsɤ1te4, tsɤ4kɤ0, tsɤ4kɤ4	嘴巴 tsuei3pa1	tʰɤ1pa1, tʰɤ1pa2, xa3pa4
椅子 ji3tsi1	tɕʰi3tsi1(2), tɕʰi3tsi4	桌子 ʈʂuo1tsi1	tɤ1ʈʂai4
拿走 na2tsou3	na2tsou3		
single C & glide 3 (.07)			
水哥 ʂuei3kɤ1	ɻuei3ka1	老公 lau3kʊŋ1	ɻau3kɤ1(4), ɻau3kɤ4, ɻau3ke1, lə3ka1
这里 ʈʂɤ4li3	tɤ2ji2		
CVCVCV 3 (.07)			
都在这 tou1tsai4ʈʂɤ4	tou1ta4tu4		
对不对 tui4pu0tui4	tu1pu0tu4		
小桌子 ɕiau3ʈʂuo1tsi0	ɕiau3tu1tu0		

Didi produced 32 different words spontaneously and 41 distinct variants. Four structures account for over 10% of the total variants each: 24% in CV, 11% in CVV, 30% in C_1VC_1V and 11% in the disyllabic variegation. The most used pattern is C_1VC_1V (30%), and the next most used one is CV (24%). Looking more closely at the two most used structures, we can see that more than half of the words are adapted using the CV structure: disyllabic words are turned into CV patterns and diphthongal monosyllables are turned into a single-vowel

monosyllables. Almost half the variants in the C_1VC_1V pattern are adapted: diphthongal disyllables are reduced to monophthongal disyllables and nasal codas are omitted to arrive at the C_1VC_1V shape. Didi makes fair use of diphthongs in monosyllables, but not in disyllables (less than 10%). Didi produces one CVC form (waŋ4, onomatopoeia) and few variegated disyllables, which all match the target structure.

Table 3.7c Xinyu's words at 25wp (18 months), ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 7 (.16)			
马 ma3	ma4(3)	奶奶 nai3nai0	nɛ1
拿 na2	na4	(上)班 paɚ1	pɤ1
沙 ʂa1	ta1(2), ta4		
这 ʈʂɤ4	tɤ4(4), ta1		
猪 ʈʂu1	tsɻ1		
CVV 7 (.16)			
美 mei3	mei4	这个 ʈʂɤ4kɤ0	tiu4(3), tou4
消失又出现 mau2	mau2	兔兔 t^hu4t^hu0	t^hei1
bye	pai1, pei2, pai2-4		
no	nuo4, nəʊ4(6)		
给 kei3	kei4; kei2		
C_1VC_1V 10 (.22)			

妈妈 ma1ma0	ma1ma0(5)	奶奶 nai3nai0	nɛ1nɛ0(6), nɛ1nɛ1
爸爸 pa4pa0	pa4pa0(4)	拜拜 pai1pai2	pa1pa0
啪啪 p^ha4p^ha4	p^ha4p^ha0	饼干 piŋ3kɛn1	ka1ka0
弟弟 ti4ti0	ti4ti0(2)		
兔兔 t^hu4t^hu0	$t^hu1tɛ0$		
哥哥 kɤ1kɤ0	kɤ1kɤ0, kɔ1kɤ0		
拉拉 la1la0	la1la0(6), la1la4, na1la0		

C_1VVC_1VV, $C_1V_0VC_1V$ 6 (.13)

杯杯 pei1pei1	pei1pei2, pei1pei0, pi1pei1(2)	饼干 piŋ3kɛn1	pei4pei0, pei1pei1(2)
拜拜 pai1pai2	pai1pei2		
妹妹 mei4mei0	mei4mei0(2)		
猫猫 mau1mau0	mau1mau1		
爷爷 jiɛ2jiɛ0	jiɛ1jiɛ0(9)		

C_1VC_2V, $C_1V_0VC_2VV_0$ 6 (.13)

大狗 ta4kou3	ta4kuo3	姐姐 tɕiɛ3tɕiɛ0	$t^hi3tɕiɛ1$, te1tɕiɛ4
八点 pa1tiɛn3	pa1ta4(3), pa1tia4	这个 ʈʂɤ4kɤ0	$tiao4k^ɕɤ4$, tiao4kɤ0, tou4kɤ4, tiao4ku4,
这边 ʈʂɤ4piɛn1	te4pe0, te4pe4		

取下 tɕʰu3ɕia4	tɕʰi3ɕiɛ4
这个 ʈʂɤ4kɤ0	kˁu1ku4, ku4ku0 ta1kɤ1, ta4kɤ0 (2)

single C & glide 7 (.16)

动物 tʊŋ4wu4	tou2wu0	飞机 fei1tɕi1	wɤ2ji1, we1tɕi1
阿姨 a1ji2	a1ji4(12), a1ji0, a2ji0	这里 ʈʂɤ4li3	kɤ4jiɛ2, kɤ4ji2
没有 mei2jou3	we4jiɛ2		
不要 pu2jau4	we2jiɛ4		
我要 wɤ3jau4	wɤ1jiɛ1 (2)		

CVCVCV 2 (.04)

在这里 tsai4ʈʂɤ4li3	ti1kɤ4ji0
Emily	a1mə0ni4, a1mu4ni0(3), a1mu4li0

Xinyu produced 39 different words spontaneously and 45 variant forms. Unlike Keke and Didi, the occurrences of Xinyu's forms are roughly equally divided across structures: disyllable CVCV patterns (10, or 22%) are slightly more frequent, other patterns like CV (7, or 16%), CVV (7, or 16%), reduplication (6, or 13%), variegation (6, or 13%) and single C with glide (7, or 16%) appear with similar frequency. All her prosodic structures mainly contain selected forms. The forms for /ba1tiɛn3/ and /ʈʂɤ4piɛn1/ in variegated disyllables are considered selected here because Xinyu is not yet producing codas.

Table 3.7d Yiyi's words at 25wp (18 months), ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 13 (.28)			
不 pu4	pu4(6)	没 mei2	mɤ2(9)
我 wɤ3	wɤ2	哞 mou1/mɤɔ1	mu1, mu4
拿 na2	na2, nɛ2	给 kei3	kɤ3, kɤ2
马 ma3	ma2(3)	写 ɕiɛ3	ɕi2
摸 muo1	mɤ1(5)	夹(上)tɕia2	ta2
车 ʈʂʰɤ1	tɕʰi1, tɕʰi4		
擦 tsʰa1	tsʰa1		
喝 xɤ1	xɤ1		
CVV 9 (.20)			
好 xau3	xau2	咚 tʊŋ4	tiu4(2)
没 mei2	mau2(4), mou2		
哞 mou1/mɤɔ1	mɤɔ1		
摸 muo1	mou1(2)		
牛 niu2	niu2		

倒 tao3 tai2

bye bai4

two t^hou2, $t^h\gamma u2$

CVC 5 (.11)

本 pən3 pəŋ2

灯 təŋ1 tɛn1

疼 t^həŋ2 t^hɛn1

拼 p^hin1 t^hɛn1, ts^hɛn1

珊 ʂan1 ɕɛn2, ɕɛn1

C_1VC_1V, C_1VCC_1VC 5 (.11)

妈妈 ma1ma0 ma1ma1, ma1ma0

哞哞 mou1mou1/ mɤə̆1mɤə̆1 mu1mu0

娃娃 wa2wa0 wa1wa0

拿拿 na2na0 na1na0(2)

蛋蛋 tɛn4tɛn0 tɛn4tɛn4, tɛn1tɛn4

$C_1VV_0C_1VV$ 2 (.04)

臭臭 ʈʂ^{h}ou4ʈʂ^{h}ou0 tɕ^{h}i1tɕ^{h}ou4

拜拜 pai1pai2 pai3pai2

C_1VCC_2V 3 (.07)

汤姆 taŋ1mu3	təŋ1na1	掉了 tiau4lə0	tɛn2nə4
		打你 ta2ni3	tɛn2ne4

single C & glide/glottal 7 (.15)

好吗 xao3ma0	xa3wa4
爱你 ai4ni3	ai1ni2
没有 mei2jou3	me2ja0
哎呀 ai2ja0	ai2ja0, ai2ja2, ai2ja4, ai4ja0, ai2jiɛ1
奥呦 au1jou4	au1jou4
hello	xa1ɹou2(4), xa2ɹou1(2), xa4ɹou1
uh-oh	o1wu2

CVCVCV 2 (.04)

没有啊	me2ja4a0
打打呀	ta4ta2ja0

Yiyi produced 38 different words spontaneously and a total of 46 distinct variants. She makes use of two high-frequency structures in her word production, CV (28%) and CVV (20%). Yiyi is the one child who produces CVC (nasal coda) to the 10% criterion at this developmental point. Her monosyllabic structures together account for 59% of her variants, two third of overall variants. There are some adapted forms in CV structure, almost all selected forms in CVV and CVC. If we combine C_1VCC_1VC (only one form) and C_1VC_1V, the resulting structure accounts for 11% of all variants, which is her only disyllabic pattern to reach criteria. The words Yiyi produced with nasal coda are both in monosyllables and syllables, although there are not enough words to posit a separate structure in disyllables. Two out of three variants are adapted that coda nasals replacing a diphthong ([tɛn2nə4] for

/tiau4lə/ ‘drop’) or monophthong ([tɛn2ne4] for /ta2ni3/ ‘hit you’). Her mastery of onset nasal may give her the attraction to coda nasal production. Yiyi did not produce variegated disyllabic word structures to the 10% criterion.

Table 3.7e Shi’s words at 25wp (18 months), ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 18 (.38)			
马 ma3	ma3, ma4	猫 mau1	mɤ1(2), mɤ4(5), mu4
怕 pʰa4	pa4(3)	鞋 ɕiɛ2	pa2, ɻe4
大 ta4	ta4	鱼 y2	βu1, βu4, ʈʂu4
是 ʂi4	ʂe4	对 tuei4	tɕi4(4), te4
鸡 tɕi1	tɕi1(5), tʂi1(5), tɕi4(3), tʂi4(3)	取 tɕʰio3	tɕi2
去 tɕʰy4	tɕi4(2)	手 ʂou3	tɕɛ4
猪 ʈʂu1	tɕɛ1	弟弟 ti4ti0	tɕi4(4)
车 ʈʂʰɤ1	tɕe1(2)	哥哥 kɤ1kɤ0	kˤɤ1(4)
蛋 tɛn4	ta4(4)	袜子 wa4tsi0	wa4 (4)
CVV, GVV 5 (.10)			
乖 kuai1	kuai1		
那 naə4	naə4		

棒 paŋ4	pau4
蛋 tɛn4	tai4
汪汪 waŋ1waŋ1	wou4

C_1VC_1V 16 (.33)

妈妈 ma1ma0	ma1ma0(6), ma1ma4(3),	奶奶 nai3nai0	na1na4
鸭鸭 ja1ja0	ja1ja0(2), ja2ja0(3), ja4ja0, ɻa1ɻa0, ɻa2ɻa0	抱抱 pau4pau0	pa1pa4, pa1pa1
饽饽 pɤ1pɤ0	pɤ1pi4, pʶɤ1pʶe1, pʶɤ1 pʶɤ0	宝宝 pau3pau0	pa3pa4
哥哥 kɤ1kɤ0	kɤ1kɤ4	姐姐 tɕiɛ3tɕiɛ0	tɛ1tɛ4, tɛ4tɛ4
娃娃 wa2wa0	wa2wa0	西瓜 ɕi1kua1	pʶɤ1pʶɤ1
汪汪 waŋ1waŋ1	wa1wa4, wa1wa1	八点 pa1tiɛn3	pa1pa0
弟弟 ti4ti0	tɕi1tɕi0, tɕi4tɕi0, tʂi4tʂi0	马 ma3	ma2ma4(2), ma3ma1, ma2ma2
姐姐 tɕiɛ3tɕiɛ0	tɕɛ1tɕi4, tɕɛ2tɕi4, tɕɛ3tɕi4	毛毛虫 mau2mau0(tʂʰʊŋ2)	mɤ2me4

$C_1VV_0C_1VV$ 5 (.10)

汪汪 waŋ1waŋ1	wa1wou4, wou4wou4	笨蛋 pən4tɛn4	pai4pai4
奶奶 nai3nai0	nai1nai0, nai1nai1, nai1nai4(2)		
抱抱 pau4pau0	pau4pau1, pau1pau4		

宝贝 pau3pei4	pa3pai1,pa3pai4,pa1pei4, pɤ1pei1, pei4pei4		
C_1VC_2V 3 (.06)			
不大 pu2ta4	pu2ta4	今天 tɕin1tʰiɛn1	tɕi1te1
		这是 ʈʂɤ4ʂi4	tɕi4ɕi0
single C & glide/glottal 1 (.02)			
		好饿 xao3ɤ4	xo1aɚ4

Shi was 17 months old at her 25wp. She produced 39 words in this session, with 48 distinct variants. She has a strong preference for two structures at this point, CV (38%) and CVCV (34%). These two patterns together account for fully 71% of Shi's word variants (or 34 out of 48); she either selects carefully having good matches to the target words or adapts some words to the structure by reducing diphthongs and omitting nasal codas. The 29% remaining forms are divided between monosyllable with diphthong CVV and disyllable with diphthong CVV_0CVV, as well as C_1VC_2V. Two of these three structures account for as much as 10% of the variants, but not variegation. Unlike Yiyi, Shi did not produce any nasal codas, although she did make high use of onset nasal.

Two children whose data were reported in the 4wp analysis left (due to their fathers' job relocation) and did not reach the 25wp within the recording period: Appendix IVa and IVb. Zuozuo produced 21 different words spontaneously and 23 variant forms at age 1;3. Yifan produced 20 different words spontaneously and 22 variant forms at age 1;4.

The general characteristics across Mandarin children will now be provided. First, an overview of prosodic structures of five Mandarin-learning children at the 25-word point is given in Table 3.8. The variant forms occurring in each distinct structure are given, out of total variants. Second, a higher criterion – 20% of the prosodic structures used by five Mandarin children – is applied and illustrated in Table 3.9 to provide a basis for cross-linguistic comparison. The children are ordered within structures from most to least proportionate use (following Vihman, 2019).

Table 3.8. Prosodic structures in Mandarin children at 25wp, ordered by proportion of disyllables

	Monosyllables			Disyllables				
	CV	CVV	CVC	C_1VC_1V	$C_1VV_0C_1V_0V$	C_1VC_2V	SingleC	CVCVCV
Keke	21% (9/44)	14% (6/44)		25% (11/44)	16% (7/44)	11% (5/44)	14% (6/44)	
Xinyu	16% (7/45)	16% (7/45)		22% (10/45)	13% (6/45)	13% (6/45)	16% (7/45)	4% (2/45)
Didi	22% (9/41)	10% (4/41)	2% (1/41)	32% (13/41)	7% (3/41)	12% (5/41)	7% (3/41)	7% (3/41)
Shi	38% (18/48)	10% (5/48)		33% (16/48)	10% (5/48)	6% (3/48)	2% (1/48)	
Yiyi	28% (13/46)	20% (9/46)	11% (5/46)	11% (5/46)	7% (3/46)	7% (3/46)	15% (7/46)	4% (2/46)
Mean	25%	14%	7%	25%	11%	10%	11%	5%
Sd	8.4%	4.2%	6.4%	8.9%	3.9%	3.1%	6.1%	1.7%

Table 3.9. Prosodic structures with 20% criterial use by Mandarin children at 25wp, ordered by proportionate within structures.

structure	child name	proportion of all variants
CV Mean .27	Shi	.38 (18/48)
	Yiyi	.28 (13/46)
	Didi	.22 (9/41)
	Keke	.21 (9/44)

CVV	Yiyi	.20 (9/46)
C_1VC_1V Mean .28	Shi	.33 (16/48)
	Didi	.32 (13/41)
	Keke	.25 (11/44)
	Xinyu	.22 (10/45)

The general tendencies in Mandarin children's prosodic structures can be observed from Table 3.8 and Table 3.9. The words produced at this point are still mostly restricted to monosyllables and disyllables, with a few longer vocalizations. The simplest phonological structures, CV and C_1VC_1V, are used over 20% by four children each. These two structures recall the first words. Furthermore, Mandarin children go beyond simple structures to more challenging forms. Diphthongs in monosyllabic words (CVV) emerge at this point. Although only one child makes critical use of CVV (20%), the other four children also use it if we consider the 10% criteria (see 3.9 for all children's use). Diphthongs in disyllabic words are also detected in three children who use over 10%. These are not typical of first words and thus may indicate developmental advance. We still see the common challenge of handling codas and consonantal variegation. Codas occur in just two of the five children's data (e.g., Didi [waŋ4] 'woof'; Yiyi [tɛn1] 'light'). Only one of them makes relatively high use (over 10%). Consonantal variegation is used to a greater or lesser extent by five children, but none of them reaches the 20% criterion, and three children make about 10 use %.

3.3.5 Cross-linguistic comparison for prosodic structure at the 25-word point

We now compare the prosodic structures of Mandarin children to those of the various language groups in Vihman (2019) to see if and how ambient languages shape the structures children use (Table 3.10). For comparison, the criteria used in Vihman (2019) are followed, that is, some more narrowly defined structures are included, such as Reduplication and Consonant Harmony.

Table 3.10 Numbers of children making 20% criterial use of each structure in Mandarin and 7 language groups adapted from Vihman (2019).

Total children	Mandarin 5	US English 6	UK English 8	Estonian 5	Finnish 5	French 4	Italian 7	Welsh 4
CVV_0	4	6	6	3	0	3	0	2
CVV	1	1	0	0	1	n/a	0	0
CVC	0	3	5	4	0	1	0	2
CH-mono	0	0	0	1	0	0	0	1
VCV_0	0	0	0	1	4	1	2	1
CH-di	0	2	5	3	5	3	4	0
REDUP	4	0	0	0	0	1	0	0
CVCCV	0	n/a	n/a	0	0	n/a	5	n/a
$C_1VC_2VC_0$	0	2	6	3	0	2	7	1
CVCVC	0	0	1	0	0	0	0	0
longer	0	0	0	0	0	0	1	0

Table 3.10 shows the cross-linguistic similarity between Mandarin and six European languages. The only comparable structures are CV and CVCV in Mandarin; none of the remaining structures reach 20% criterion except for one child's use of CVV. The most common structures in Mandarin children – CV and CVCV – are also found in six or seven of the other seven groups. The structures of Mandarin are simpler compared with the European languages (as introduced in chapter 1); accordingly, the other patterns found in those languages are not used to criterion by the Mandarin children.

There are sharp differences: Disyllabic consonant harmony accounts for often-used structures in all European groups (except Welsh), while the use of this pattern is not seen in Mandarin. Instead, reduplication occurs as a high-use structure on its own for four out of five Mandarin children but in only one European (French) child. Diphthongs seem to be a characteristic of Mandarin children. The CVV structure is found in only one US English child and one Finnish

child among all European groups. Although it is also used to the 20% criterion by only a single Mandarin child, Table 3.11 brings out the difference more if we relax the criterion to 10% use: all five children favour this diphthongal monosyllabic structure. Three out of five children also use the diphthongal disyllabic structure (C_1VVC_1VV) to this lower criterion level. Recall that Mandarin has a rich inventory of diphthongs as well as triphthongs. None of the Mandarin children make frequent use of variegation, while in the other languages half or more (UK English, Estonia, French, Italian) or at least some children (US English, Welsh) use it to criterion.

Table 3.11 Prosodic structures with 10% criterial use by Mandarin children

	Keke	Didi	Xinyu	Yiyi	Shi
CV	✓	✓	✓	✓	✓
CVV	✓	✓	✓	✓	✓
CVC				✓	
C_1VC_1V	✓	✓	✓	✓	✓
$C_1V_0VC_1VV_0$	✓		✓		✓
$C_1V_0VC_2VV_0$	✓	✓	✓		
C+glide/glottal	✓		✓	✓	

It is instructive to look at variegation closely to see the cross-linguistic differences since none of the Mandarin children use this structure to 20% criteria. In comparison with European-language-learning children (except Finnish), Mandarin children are slow to attempt variegated sequences. Mandarin learners attempt only 34% variegated targets out of all disyllables, in contrast to Italian children, who produce the highest proportion of variegated targets (71%). Children learning the other languages attempt variegated targets to differing degrees between those extremes (Table 3.12). Despite the fact that producing two different supraglottal consonants in a single word form at this developmental point is the biggest challenge for most children in many languages, there are different ways to reduce the overall

complexity. Children learning European languages use a variety of strategies, with consonant harmony (partial reduplication, with repeated consonants, but not vowels) documented as the most widely used phonological process (Smith, 1973; Vihman & Greenlee, 1987). However, consonant harmony is not 'universal' in child phonology and is not a common process in Mandarin-learning children. Mandarin children mainly restrict their disyllabic forms either to full reduplication or to accurate production of the variegated targets. We will discuss these findings in relation to the characteristics of Mandarin input speech in a later section.

Table 3.12 Variegated targets and dominant child forms, by language (from BAAP, 2020)

Language group	N children	N disyllables / all variants	Variegated targets / all disyl. variants	Child varieg. forms / disyl. vars. targeting variegated words	Main child forms for var. targets
US English	6	0.36	0.69	0.29	Consonant harmony and 'other'
UK English	6	0.39	0.52	0.14	Consonant harmony
French	5	0.57	0.63	0.32	Consonant harmony
Italian	7	0.80	0.71	0.63	Variegation
Finnish	5	0.78	0.52	0.29	Consonant harmony
Mand-Y	5	0.52	0.34	0.53	Variegation
mean		0.59	0.56	0.39	

3.3.6 Consonant inventories and possible templates in Mandarin-acquiring children at 25-word point

We will investigate children's segmental resources, in other words, the challenge of consonant production in children's early production at the end of the single-word period. We will present a consonant inventory for each child in both onset and coda position in all forms of the words they produced (with or without match to target). Combining this with prosodic

structures which are presented earlier in the chapter, we then will see if any particular specified patterns which can be called templates can be found in any child.

Template identification

Phonological templates are seen as overuse of a particular specified pattern by a given child compared to other children learning the same language. There are two ways to identify templates: prosodic structures are specified on the basis of children over-selecting words to attempt that fit their familiar patterns or adapting difficult words to fit their patterns (Vihman, 2014). Segmental patterns like particular consonants or vowels, or consonant and vowel classes or consonant and vowel melodies are applied to 20% of variants or more (see Kehoe, 2015). Templates can only be detected when children produce around 20-50 words but fewer than 200 words (Renner, 2017). The twenty-five word points are analysed in the Mandarin children to see if some of them may have templates.

The following criteria are applied to illustrate use of consonants in our counts and tables (based on Vihman, 2019): Two child forms from different target words are provided in each of two positions (onset and coda), where available. For more than two uses of a consonant, we indicate the total number of words in column T. We count no more than one use of a consonant for a single target word but repeat words as needed for different consonant uses. For example, Keke's production [p] in [pa4pa0] and [pa1pa2] are listed; there are eight uses in total, including these two. We are aiming to find the favoured segments and individual patterns of each child.

Keke

Keke was 17 months old at her 25wp. She produced 37 words in this session with 44 variants. Keke's consonant inventory for each sample is shown in Table 3.13a.

Table 3.13a. Keke's consonant inventory at 17 months. Bold face means target consonants are matched in at least two child forms.

target C (N=9)	onset	T	coda	T

p	pa4pa0 [pa4pa0]	pai1pai2[pa1pa2]	7
t^h	ɕiɛ4ɕiɛ0 [t^hi1tɕiɛ0]		
t	tiau4tiau0 [tau4tau0]	pu2tiau4[pu1tau1]	4
k^h	k^han4 [k^hɤ4]	k^hai1 [k^ho4]	8
k	kai4 [kɤə4]	kua1tsi3 [ku1si0]	3
m	ma1 [ma1]	mia1mia1 [mia4mia4]	3
n	nai1nai1 [nə1nə0]		
f	yi1fu2 [fu1]	lao2xu3 [lou1fu4]	4
	xui4 [fei4]	wo3xui4 [en1fei4]	
s	gua1tsi3 [gu1si0]		
ʂ	ʂou3 [ʂou2]		
tʂ	tʂɤ4 [tʂɤ4]		
ɕ	ɕiɛ4ɕiɛ0 [ɕiɛ1ɕiɛ4]	ɕio1ɕi0 [ɕio1ɕi4]	3
tɕ^h	ɕiɛ4ɕiɛ0 [kɤ1tɕ^{h}iɛ]		
l	lao2xu3 [lou1fu4]	hello [a1lou1]	
ɻ	ai1ja0 [ai4ɻa0]		
w	wa4 [wao4]	pao4wo3 [pa4we1]	3
j	ju4t^hou2 [ju2tou2]	hiya [ai2ja1]	

Keke's consonant inventory chart

p		t		k
		(t^h)		k^h
			(ʈʂ)	
				(tɕ^h)
m		(n)		
	(f)	(s)	(ʂ)	ɕ
w			(ɻ)	j
		l		

Keke explores a variety of consonants and produces them to similar extents. She produced nine onset consonants as match to target in more than one word-type: the stops [p, t, k, k^h], the nasal [m], the fricative [ɕ], the lateral [l] and glides [w, j]. The table reveals Keke's preference for producing the stops [p] and [k^h]. She also makes frequent use of the stop [t] and the fricative [f], but most of her uses of [f] are as a consistent phonetic substitute for [x] rather than as a match to the target. Keke is one of only two children who include unaspirated obstruents in her repertoire.

Table 3.13b. Keke's preferred pattern: 27% of total variants

prosodic structure	selected words target	child form	adapted words target	child form
CV	盖 kai4	kɤ1	手 ʂou3	k^ho2
	开 k^hai1	k^ho1, k^ho4	这 ʈʂɤ4	kɤ4

	看 kʰan4	kʰɤ4		
CVCV	可可 kʰɤ2kʰɤ3	kʰɤ2kʰɤ3, kʰɤ1kʰɤ4		
	盖盖 kai4kai0	kɤ1kɤ1(2), kɤ1kɤ0		
$C_1VC_2VV_0$	瓜子 kua1tsi3	ku1si0	盒子 xɤ2tsi0	kʰɤ3kɤ1
			谢谢 ɕiɛ4ɕiɛ0	kʰɤ1tɕʰiɛ4
			鞋鞋 ɕiɛ2ɕiɛ0	kʰɤ2ɕiɛ2, kʰɤ2ɕi2

Keke has two most-used structures (as noted earlier): CV (9, or 21%) and CVCV (11, or 25%). Some of the variants are unlike their targets but similar to one another. If we look closely at Keke's favorite consonant [kʰ]/[k], we can see that the two most used structures all have a salient /k/-initial pattern: six out of eleven are selected /k/-initial words and five are other words adapted to this output form (Table 4.3.5b). The preference for initial velar [kʰ] / [k] with back vowel in monosyllables and as a part in disyllables. If we analyse this as a potential template pattern, we find they account for 27% of Keke's total word variants.
The effect of favouring velar back vowel association is quite prominent in Keke's words. This association, which is one of the three CV co-occurrences proposed by Davis and her colleagues (Davis, MacNeilage & Matyear 2002), is a general characterization of intracyclic properties of early vocalizations. A study investigating the first 50 words of 23 children in four different language communities also confirmed the velar-back vowel co-occurrence constraint (Vihman, 1992).

Didi

Didi explores a large number of consonants at 17 months (Table 3.14a). Only five onset consonants match their targets: the stops [p, t, k], the nasal [m], and the affricate [ts]. It is clear that Didi makes heavy use of the alveolar consonant [t], which is three times more frequent than the next most-used consonants [p], [k] and [m]. We can see that aspirated and unaspirated obstruents are not well distinguished; there are not many unaspirated obstruents and only a few are accurately produced.

Table 3.14a Didi's consonant inventory at 17 months. **Bold face** means target consonants are matched in at least two child forms.

target C (N=5)	onset		T	coda	T
p	**pao3pao0 [pa3pa1]**	**pai1pai2 [pai1pai2]**	5		
t^h	ʈʂɤ4 [t^hɤ4]	tɕiɛ3tɕiɛ0 [t^hɛ3t^hɛ0]	3		
t	**ti4 [ti4]**	**ta2ta0 [ta2ta0]**	15		
k^h	go [k^hou1]				
k	**ka1ka1 [ka1]**	**kei3 [kei2]**	5		
m	**ma1 [ma1]**	**mao1 [mao1]**	5		
n	na2tsou3 [na2tsou3]				
x	tsui3ba1 [xa3ba4]				
ts	**ʈʂuo1tsi1[tse1tsi0]**	**ji3tsi1 [tɕ^{h}i3tsi1]**	3		
ʈʂ^h	ʈʂ^{h}i1 [ʈʂ^{h}i1]				
ʈʂ	ʈʂuo1[ʈʂu1]	ʈʂuo1tsi1 [tou1ʈʂai4]			
tɕ^h	ji3tsi1[tɕ^{h}i3tsi1]	tɕiɛ3tɕiɛ0 [tɕ^{h}i3tɕiɛ1]			
tɕ	tɕiɛ3tɕiɛ0 [tɕiɛ3tɕiɛ0]				
l	lao3gʊŋ1 [lə3ga1]				
ɻ	ʂui3gɤ1 [ɻui3ga1]	lao3gʊŋ1 [ɻo3gɤ1]			
w	waŋ4[waŋ4]				
ŋ				waŋ4 [waŋ4]	

Didi's consonant inventory chart

p	**t**			**k**
	(t^h)			(k^h)
	ts	(ʈʂ)	(tɕ)	
		($ʈʂ^h$)	($tɕ^h$)	
m	(n)			(ŋ)
				(x)
(w)		(ɻ)	(j)	
	(l)			

Didi displayed frequent use of alveolar /t/ either as a reduplicated pattern CtVCtV or as a /t/-initial pattern in most used structures (see Table 4). Eight out of the 13 words in his most-used CVCV structure (13, or 32%) have tVtV - [t] onsets with a range of different vowels, high and low, front, central and back. This includes both selected and adapted words. Almost half (4 out of 9) of his next-most-used CV structure (9, or 22%) reflect a t-initial preference. Although the proportion of the variegated structure among Didi's prosodic structures is small (12%), the evidence of preferred [tɤ] pattern is relatively strong, as more than half of the word variants included are adapted to the [tɤ] pattern. These patterns together account for 39% of his word variants.

Table 3.14b. Didi's preferred pattern: 39% of total variants

prosodic structure	selected words target	child form	adapted words target	child form
CV	弟弟 ti4ti0	ti4(3), ti2	车 $ʈʂ^hɤ1$	$t^hɤ4$

	滴滴 ti1ti1	ti1	这 ʈʂɤ4	tɤ1(2)
C_tVC_tV	打打 ta2ta0	ta2ta0(2), ta4ta0	姐姐 tɕiɛ3tɕiɛ0	tɛ3tɛ0(2)
	滴滴 ti1ti1	te1te1(2)	车车 ʈʂʰɤ1ʈʂʰɤ1	tɤ1tɤ4(2)
	弟弟 ti4ti0	ti4ti0, ti1ti1(2)	走走 tsou2tsou0	tu1tu0
	痛痛 tʰʊŋ4tʰʊŋ4	tu4tu4, tu4tu0(2)	在这 tsai4ʈʂɤ4	ta3tu4, tu2tu4(2) tu4tu4, tu4tu0
C_1VC_2VV	都在 tou1tsai4	tɤ1ʈʂai4	这个 ʈʂɤ4kɤ4	tɤ1kɤ4
			嘴巴 tsui3pa1	tʰɤ1pa1, tʰɤ1pa2
			这里 ʈʂɤ4li3	tɤ2ji2
CVCVCV	都在这 tou1tsai4ʈʂɤ4	tu1ta4tu4	小桌子 ɕiao3ʈʂuo1tsi0	ɕiao3tu1tu0
	对不对 tui4pu0tui4	tu1pu0tu4		

Xinyu

Xinyu's inventory is quite restricted. She produces thirteen consonants and nine of them match the targets: the stops [p, t, k], the nasal [m, n], the affricate [tɕ], the lateral [l] and glides [w, j]. [tɕʰ] and [ɕ] are produced accurately in only a single word and so are not included in her inventory. She does not attempt words which are far beyond her phonetic skills. [p], [t] and [k] are mainly for unaspirated targets. Nasals [m] and [n] also occur only in onset position.

Table 3.15a Xinyu's consonant inventory at 17 months. **Bold face** means target consonants are matched in at least two child forms.

target C (N=9)	onset		T	coda	T
p	**paɚ1 [pɤ1]**	**pa4pa0 [pa4pa0]**	7		
t^h	tɕiɛ1tɕiɛ0 [t^hi1tɕiɛ1]	t^hu4t^hu0 [t^hei1]			
t	**ʈʂɤ4 [tɤ4]**	**ʂa1 [ta1]**			
k	**kɤ1kɤ0 [kɤ1kɤ0]**	**piŋ3kɛn1 [ka1ka0]**	7		
m	**ma1 [ma4]**	**mao1mao0 [mao1mao1]**	6		
n	**na2 [na4]**	**nai3nai0 [nɛ1nɛ0]**	5		
ts	ʈʂu1[tsɻ1]				
tɕ	**tɕiɛ3tɕiɛ0 [t^hi3tɕiɛ1]**	**fei1tɕi1 [we1tɕi1]**			
$tɕ^h$	$tɕ^h$y3ɕia4 [$tɕ^h$i3ɕiɛ4]				
ɕ	$tɕ^h$y3ɕia4 [$tɕ^h$i3ɕiɛ4]				
l	**la1la0 [la1la0]**	**Emily [a1mu4li0]**			
w	**wɤ3jao4 [wɤ1jia1]**	**tʊŋ4wu0 [tou2wu0]**	5		
j	**jiɛ2jiɛ0 [jiɛ1jiɛ0]**	**a1ji2 [a1ji4]**	8		

Xinyu's consonant inventory chart

p	t		k
	(t^h)		
		ʈʂ	tɕ
			($tɕ^h$)
m	n		
			(ɛ)
w			j
	l		

Note that Xinyu produces the palatal glide /j/, word-final /i/ or a front diphthong nucleus (as match to target or not) in more words than any other (see Table 3.15b). This thus appears to be a distinct pattern. The CVGLIDEV structure is Xinyu's second-most-used structure. Although the [ji] pattern accounts for only 18% of Xinyu's total variants, almost all the word variants (6 out of 7) included in her CVGLIDEV structure take this pattern.

Table 3.15b Xinyu's preferred pattern 18% of total variants

prosodic structure	selected words target	child form	adapted words target	child form
(C)VC_jVV_0	a1ji2	a1ji4(12), a1ji0	fei1tɕi1	wɤ2ji1
	mei2jou3	we4jiɛ2	ʈʂɤ4li3	kɤ4jiɛ2
	pu2jao4	we2jiɛ4		
	wɤ3jao4	wɤ1jia1, wɤ1jiɛ1		
C_jVVC_jVV	jiɛ2jiɛ0	jiɛ1jiɛ0		
CVCVCV				

Yiyi

Yiyi has eight onset consonants as match to target in more than one word type: the stops [p, t, tʰ], the nasals [m, n], velar fricative [x] and glides [w, j]. Surprisingly, unlike the other four children, who all have /p, t, k/ in their consonant inventory, she uses [k] in just one word. However, Yiyi preferred to produce nasals in both onset and coda position, and she is the only child who produced nasal codas to any great extent (we see a single variant [waŋ4] in Didi's production). The other favored sound is [t]. Yiyi produced more aspiration compared to the others, although her use does not always match the target.

Table 3.16a Yiyi's consonant inventory at 17 months. **Bold face** means target consonants are matched in at least two child forms.

target C (N=8)	onset		T	coda		T
p	**pu4 [pu4]**	**pai1pai2 [pai3pai2]**	4			
tʰ	**tʰəŋ2 [tʰən1]**	**two [tʰɤu2]**	3			
t	**tao3 [tai2]**	**tɛn4tɛn0 [tɛn4tɛn4]**	7			
k	kei3 [kei2]					
m	**mɤ1 [mɤ1]**	**mou1 [mu1]**	5			
n	**na2 [na2]**	**niu2 [niu2]**	5	**pʰin1 [tʰɛn1]**	**ʂan1[ɕɛn1]**	6
ŋ				**taŋ1mu3 [təŋ1na1]**	pən3 [pəŋ2]	
ɕ	**ɕiɛ3 [ɕi2]**	ʂan1[ɕɛn1]				
x	**xɤ1 [xɤ1]**	**xao3 [xao2]**	3			
tsʰ	tsʰa1[tsʰa1]	pʰin1[tsʰən1]				
tɕʰ	tʂʰɤ1 [tɕʰi1]	tʂʰou4tʂʰou0				

[tɕʰi1tɕʰou4]

ɻ hello [xa1ɻou2]

w **wo3 [wɤ2]** **wa2wa0 [wa1wa0]** 4

j **ai2ja0 [ai2ja0]** **au1jou4 [au1jou4]**

Yiyi's consonant inventory

p	**t**			(k)
	tʰ			
	(tsʰ)		(tɕʰ)	
m	**n**			(ŋ)
			(ɕ)	**x**
w		(ɻ)	**j**	

Table 3.16b Yiyi's preferred pattern: 20% of total variants

prosodic structure	selected words target	child form	adapted words target	child form
CVC	pʰin1	tʰɛn1, tsʰɛn1	təŋ1	tɛn1
	ʂan1	ɕɛn2, ɕɛn1	tʰəŋ2	tʰɛn1
			pən3	pəŋ2
CVCCV(C)	tɛn4tɛn0	tɛn4tɛn4, tɛn1tɛn4	ta2ni3	tɛn2ne4
	taŋ1mu3	təŋ1na1	tiao4lə0	tɛn2nə4

A nasal coda pattern was found in Yiyi's production in two forms: [ɛn] and [əŋ]. All these patterns constitute CVC monosyllables and variegated disyllabic structures, although these two structures are not the most used structures. Note that Yiyi is the only child among the Mandarin children who has an emerging CVC structure. This specific pattern may be considered a 'template' and accounts for 20% of Yiyi's variants. It is important to recall that Mandarin has only nasal codas. Yiyi is the only child who also uses large numbers of word-initial nasals; therefore the initial nasals could be a hook that attracts the production of coda nasals. This supports Pierrehumbert (2003) "an outcome which represents one phoneme in one context may represent a different phoneme in a different context". Here, the one articulatory gesture connects both the onset and coda nasal. Experience producing onset nasals may support learning coda nasals.

Shi

Table 3.17a Shi's consonant inventory at 17 months. **Bold face** means target consonants are matched in at least two child forms.

target C (N=7)	onset		T	coda	T
p	pʰa4 [pa4]	paŋ4 [pao4]	9		
t	ta4 [ta4]	tɛn4 [ta4]	5		
k	kɤ1 [kˤɤ1]	kuai1 [kuai1]			
m	ma3 [ma3]	mao1 [mɤ1]	4		
n	naɚ4 [naɚ4]	nai3nai0 [nai1nai0]			
ʂ	ʂi4 [ʂe4]				
x	xao3ɤ4 [xo1aɚ4]				
[β]	y2 [βu1]				
ʈʂ	y2 [ʈʂu4]				

tɕ	tɕi1 [tɕi1]	tɕiɛ3tɕiɛ0 [tɕɛ1tɕi4]	11
ɻ	ja1ja0 [ɻa1ɻa0]		
w	wa4 [wa4]	waŋ4waŋ4 [wa1wa4]	
j	ja1ja0 [ja1ja0]		

Shi's consonant inventory chart

	p	t		k
		(tʂ)	tɕ	
	m	n		
	[β]	(ʂ)		(x)
	w	(ɻ)	(j)	

Shi is quite a cautious child who does not do much exploration. She produced 13 consonants and ten of them matched their targets. The consonants [ʂ] [x] and [j] are produced accurately in only a single word each and so are not included in her inventory. Thus she has the stops [p, t, k], the nasal [m], the affricate [tɕ] and glides [w]. She does not produce any aspirated consonants.

Table 3.17b Shi's preferred pattern: 25% of total variants

prosodic structure	selected words target	child form	adapted words target	child form
CV	鸡 tɕi1	tɕi1(5), tʂi1(5), tɕi4(3), tʂi4(3)	弟弟 ti4ti0	tɕi4(4)
			对 tui4	tɕi4(4)
	取 tɕʰio3	tɕi2	去 tɕʰy4	tɕi4(2)
			猪 tʂu1	tɕɛ1
			车 tʂʰɤ1	tɕe1(2)
			手 ʂou3	tɕɛ4
$C_{tɕ}VC_{tɕ}V$	姐姐 tɕiɛ3tɕiɛ0	tɕɛ1tɕi4, tɕɛ2tɕi4, tɕɛ3tɕi4	弟弟 ti4ti0	tɕi1tɕi0, tɕi4tɕi0, tʂi4tʂi0
C_1VC_2V	今天 tɕin1tʰiɛn1	tɕi1tɛ1	这是 tʂɤ4ʂi4	tɕi4ɕi0

Shi displayed a /tɕ/-initial pattern. CV (38%) is Shi's most used structure. Of the 18 variants in this structure, eight forms begin with the palatal affricate [tɕ], with high ([i]) or mid front ([e, ɛ]) vowels. There is strong evidence of template use in this case, as almost all the word variants are adapted. This pattern is also found in reduplication and variegation structures. Shi's templatic pattern makes up 25% of all her variants.

3.3.7 General tendenies in consonant and template use in Mandarin-acquiring children at 25-word point

General tendencies can be seen in the consonant inventory. Some consonants are present in all of these Mandarin children's phonetic inventories by the end of the single-word stage

(25wp) : the stops /p, t, k/, the nasal /m/ and the glide /w/. If palatal [j] is classified as a coronal, and the glide [w] is classified as a labial, the order of frequency of use in consonant place was coronal (47%), labial (37%) and dorsal (16%). For consonant manner, the order of frequency of use was stops (43%), fricatives and affricates (23%) nasals (18%) and glides (16%). All the children produce a high proportion of labials and stops in this period. The use of fricative/affricative is higher than nasal. Every child produced at least some fricatives and affricates; fricative [ɕ] and affricate [tɕ] are the most common. They expand single word with fricatives or affricates in relevant routine contexts (e.g. /tɕiɛ3tɕiɛ/ 'sister', /ɕiɛ4ɕiɛ0/ 'thanks') to more words. However, unaspirated obstruents are still hard to manage for Mandarin children at this development point, when aspirated and unaspirated obstruents are not well distinguished in production. Fricatives/affricates still substitute for stops in some children's variants and occur in coda position rather than onset for some children (e.g. Keke). In addition, vowels emerge and are mastered much earlier than consonants in many languages. However, substitutions were found in diphthong and triphthong production in these children.

Based on the analysis of Mandarin-learning children's words, heavy use of some patterns provides evidence for template at the 25-word point. Table 3.18 summarises the results. Although they are all growing up in the same language community, there are individual differences of template use among Mandarin children, that is, the children form their own preferred patterns. The combined prosodic structure and segment inventory analyses helped us to identify the template use. A child's favoured pattern sometimes includes more than one prosodic structure. Of the five children, two have prosodic templates: Xinyu makes templatic use of $CV_{Glide}V$ structure; Yiyi makes templatic use of nasal coda in both monosyllabic and disyllabic words. Three children have specific consonantal templates: They either over-select CV with a specific consonant, or they manage the challenge by constraining the first consonant of disyllabic words to their preferred consonant. Some children rely on templates more than others.

Table 3.18. Templates identified for the Mandarin children

child name	templatic pattern	structures	proportion of total
Keke	initial /k/ with back vowel	CV, CVCV C_1VC_2V	.27

Didi	/t/-initial	CV, CVCV C_1VC_2V	.39
Shi	initial /tɕ/ with front vowel	CV, CVCV C_1VC_2V	.25
Xinyu	palatal glide /j/ with /i/ or front diphthong nucleus	(C)VGLIDEV	.18
Yiyi	Nasal coda: [ɛn] or [əŋ]	CVC, CVCCV(C)	.20

Templates are an accommodation between the phonological forms of an adult target word and the child's own repertoire, presented as selection or adaptation. As introduced in chapter 1, however, different approaches identify templates differently. According to Vihman (2014), the term template refers to syllabic patterns like CV monosyllables, CVCV disyllables and VCV or CVC, or melodic patterns like consonant harmony. Kehoe (2015) also describes the use of initial consonants or specific consonant codas as template.

The Mandarin children are similar to the children in Vihman (2014): $CV_{GLIDE}V$ or CVC serve as patterns, and a number of cases of CV are a pattern. However, in some ways the children are more like Kehoe (2015)'s German- and Spanish-learning children. The case of Keke's /k/-initial, Didi's /t/-initial and Shi's /tɕ/- initial pattern recall Kehoe's findings (2015) in showing that children may have consonant-initial patterns as templates. These children either select /k/, /t/ or /tɕ/ as a default onset or use them as substitution for other words, respectively.

3.3.8 Cross linguistic comparison for consonantal resources at 25-word point

We will draw a cross-linguistic picture of consonant in terms of manners of articulation. As indicated earlier, Mandarin children produce many fricatives and affricates, although late acquisition of affrication has been reported in many languages. Data for children in other language groups are calculated based on Vihman (2019, Chapter 5) and align with Mandarin

children of 25-word point sessions. Table 3.19 is ordered by the proportion of fricatives/affricates.

Table 3.19 Comparison for consonant manners between Mandarin children and 7 European language groups

Language group (N children)	Stop	Fricative/Affricate	Nasal	Approximant
Mandarin (5)	43%	23%	18%	16%
Welsh (2)	47%	17%	23%	13%
French (2)	52%	16%	8%	25%
UK English (6)	60%	10%	19%	12%
Estonia (2)	59%	9%	23%	9%
Finish (2)	80%	9%	11%	0
US English (2)	57%	7%	20%	16%
Italian (2)	58%	2%	22%	18%

Some consonants were used a similar amount across languages. For example, stops are the dominate consonant (at least half or almost half out of all types) for all children.
The fricatives and affricates usually emerge late in English-learning children, as can be seen both from the literature and from this table, and also in children acquiring three other languages (Estonia, Finish, Italian), while Mandarin-speaking children reveal some language-specific characteristics in that their use of fricatives and affricates is the highest of the eight language groups. The fricatives /ts, ʈʂ, tɕ/, affricate /ɕ/ and one velar fricative /x/ emerge at about the same time as labial stops in Mandarin children's phonetic inventories here.

It is known that Mandarin has a rich inventory of fricatives and affricates at different places of articulation, as shown in Chapter 1. Compared with Mandarin, English has mainly anterior fricatives and lacks retroflex, palatal and velar consonants. One suggestion is that the intensive exposure might make Mandarin children better able to produce affricates and

fricatives. As children acquire the ambient language, they must learn the structures and the inventories that are specific to their target language. We will examine the Mandarin input speech in next section.

3.4 Characteristics of Mandarin input speech

In order to understand where the children's prosodic and segmental patterns come from, this chapter will first investigate the language structure and phonetic frequency in the input speech of Mandarin. We will then compare Mandarin input speech to languages that differ in certain phonological and prosodic properties to see whether cross-linguistic similarities are detectable in infant-directed speech (IDS) and to what extent cross-linguistic differences in certain properties are detectable in IDS. This section presents primary data of Mandarin input speech and aims to find out how often children hear the specific patterns, and therefore what contributes to children's early prosodic structures and segmental patterns. The phonological patterns to be investigated here are those that were reported for infant production in previous sections.

To obtain the phonetic characteristics of the input speech, mothers of 12-month-olds were transcribed. In order to follow the same criteria that we applied to children's vocalizations, we analysed the distribution of prosodic structures (length in syllables) and manners of articulation. We broke down the caregivers' words in running speech by word length: monosyllable and disyllable.

We transcribed and analysed both content and function words in all child-directed running speech. Content words included all nouns, main verbs, adjectives, adverbs, conventional interjections (/ai1jou0/ 'uh-oh'), and onomatopoeia (/miau1miau1/ 'meow', /wang4/ 'woof') as well as simple formulaic routines (/mau2/ 'peek-a-boo'). Verb particles were counted as part of the verb (/fang4ɕia4/ 'put down', /tɕʰi3lai2/ 'stand up'). Function words included articles, auxiliaries and conjunctions, all forms of the prepositions, quantifiers, question words. Sound effects, unconventional onomatopoeia (/rua4rua4/ for a dinosaur sound), and imitations of the child's production were not treated as content words. Nursery rhymes and songs were excluded. The manner of articulation for supraglottal used. Manner categories were divided into (1) stops, (2) fricatives/affricates, (3) nasals, and (4) approximates. We analysed the frequency of stops, fricatives/affricates, nasals, and approximants for eight

mothers' speech. The IDS conversations of the eight mothers contained 1620 word types and 15323 word tokens.

3.4.1 Prosodic structures in Mandarin input speech

Table 3.20 Type frequency of prosodic structures in caregivers' speech

	Monosyllables			Disyllables			
	CV	CVV	CVC	C_1VC_1V	C_1VVC_1VV CVNCVN	C_1VC_2V	SingleC
Zuozuo	17.6%	20.9%	4.4%	2.2%	9.9%	31.9%	13.2%
Yifan	10.8%	18.0%	12.6%	4.5%	11.7%	31.5%	10.8%
Keke	14.2%	19.5%	8.8%	5.3%	12.4%	29.8%	9.7%
Xinyu	17.0%	23.4%	11.7%	3.2%	12.8%	25.5%	6.4%
Didi	18.5%	21.0%	8.6%	6.2%	9.9%	29.6%	6.2%
Yiyi	17.3%	20.2%	11.5%	2.9%	11.5%	30.8%	5.8%
Shi	13.1%	21.5%	13.8%	1.5%	9.2%	34.6%	6.2%
PingAn	15.0%	22.4%	8.4%	3.7%	10.3%	34.6%	5.6%
Mean	15.4%	20.9%	10.0%	3.7%	11.0%	31.0%	8.0%
Sd	2.6%	1.7%	3.0%	1.6%	1.3%	2.9%	2.9%

Table 3.20 shows the distribution of the prosodic structures (in types) for the eight Mandarin speaking caregivers. If we first look at the mean for all structures we can see monosyllabic words with or without diphthong and variegated disyllabic words characterize the speech of all of the caregivers. For monosyllables, mean use of words with diphthong (20.9%, range 18.0% - 23.4%) was higher than words with monophthong (15.4%, range 10.8% - 18.5%). For disyllables, words with variegation dominate (31.0%): this is more than twice as much as the next most often used pattern, diphthongal reduplicated words (11%). (G)VCV and C_1VC_1V occur only rarely (8.0% and 3.7%, respectively). The prosodic structures of adult input speech are quite similar across the different speakers, as shown by the relatively low

standard deviation.

3.4.2 Manner of articulation in Mandarin input speech

We conducted an analysis of 8 mothers including all consonant manners of articulation: stops /p, pʰ, t, tʰ, k, kʰ/, fricatives /f, s, ʂ, ɕ, x/, affricate /ts, tsʰ, ʈʂ, ʈʂʰ, tɕ, tɕʰ/, nasals /m, n, ŋ/ and approximant /r, l, w, j/. For each of the five types of consonants, type frequencies at each position within a word were analysed.

Table 3.21 Distribution of consonantal manner categories in caregivers' speech (in %)

	Stop	(Af)fric	Nasal	Approximate
Zuozuo	38.0% (57/150)	34.0% (51/150)	14.7% (22/150)	13.3% (20/150)
Yifan	31.9% (52/163)	40.5% (66/163)	11.7% (19/163)	16.0% (26/163)
Keke	33.0% (58/176)	39.2% (69/176)	9.7% (17/176)	18.2% (32/149)
Xinyu	32.2% (48/149)	38.9% (58/149)	12.8% (19/149)	16.1% (24/149)
Didi	34.4% (32/93)	38.7% (36/93)	9.7% (9/93)	17.2% (16/93)
Yiyi	34.2% (53/155)	34.2% (53/155)	12.3% (19/155)	19.4% (30/155)
Shi	31.3% (57/182)	45.1% (82/182)	8.8% (16/182)	14.8% (27/182)
PingAn	32.2% (49/152)	38.8% (59/152)	14.5% (22/152)	14.5% (22/152)
Mean	33.4%	38.7%	11.7%	16.2%
Sd	1.3%	3.5%	2.1%	2.0%

Table 3.21 shows the distribution of manner of articulation categories in the speech of the eight caregivers. Mandarin input speech shows overall high use of stops (33.4%) and fricatives/affricates (38.7%). Nasals and approximates are used less in these data.

3.4.3 Cross linguistic comparison in consonants manner and prosodic structure in input speech

To gain a better picture of the phonological characteristics of input speech and a possible role for input characteristics in the association with children's production, Mandarin input speech was compared with three different European languages (English, French and Swedish), as reported in Vihman et al. (1994): See Table 3.22.

Table 3.22. Manners of articulation and length in syllables in Mandarin input Versus European language sample comparisons (in %; polysyllables are also in the count and therefore 100% total rate includes polysyllables).

		Mandarin	English	French	Swedish
Stop	M	33.4	44.5	43.2	39.0
	SD	1.3	7.1	5.6	1.8
(Af)fric	M	38.7	18.9	18.1	23.1
	SD	3.5	3.2	2.0	2.5
Nasal	M	11.7	17.6	15.0	19.1
	SD	2.1	5.3	6.2	2.0
Approx	M	16.2	19.0	23.8	18.9
	SD	2.0	3.5	6.2	2.0
Mono	M	41.6	69.2	56.3	43.8
	SD	2.4	7.7	7.4	8.7
Di	M	56.1	22.9	38.6	46.1
	SD	2.5	3.3	7.0	3.7

With respect to the phonetic profile, there are differences across the languages. Specifically, stops constitute the most preferred consonant category for all mothers in the three European languages (English, French and Swedish), but not for Mandarin; instead, fricative/affricate is the most common category for Mandarin-speaking mothers. Nasals and approximants are less used and did not differ much between languages; none of the language groups used nasals in more than 20%. As for length in syllables, monosyllables dominate English and French mothers' speech, and are more or less as frequent as disyllables in the speech of Swedish mothers. However, disyllables constitute the dominant category in Mandarin mothers' speech.

As mentioned in chapter 1, in Mandarin disyllabic words account for the highest proportion of words, 74.3%. There are not many differences in terms of structure between Mandarin ADS and IDS. The privilege of stops in Mandarin is also found in other languages and can be seen as a language-general tendency. The Mandarin preference for fricatives/affricates as well as for disyllables confirm the existence of language-specific characteristics.

3.3.4 Discussion

To explore possible factors that determine the emergence of phonological patterns during development of Mandarin children, this chapter has described adult production to infants learning Mandarin. Different frequency analyses in terms of prosodic structure and consonant inventory were conducted on the Mandarin input speech.

The input data indicate that disyllables are used more than monosyllables, in line with the characteristics of Mandarin ADS. This is in contrast with other three European languages, where monosyllables have been found to use more than disyllables in English and French and relatively equally used to disyllables in Swedish (Vihman et al., 1994).

Since Mandarin input speech has not yet been extensively documented, it is worth noting that Mandarin-speaking mothers used many words with fricatives/affricates (e.g., /tɕiɛ3tɕiɛ0/ 'sister', /ɕiɛ4ɕiɛ0/ 'thanks', /tɕʰin/ 'kiss', /ʈʂɤ4/ 'this', /ʈʂʰi1/ 'eat', /ʂu1/ 'book', /ja1tsi0/ 'duck', /tsʰao3mei2/ 'strawberry', /tou2fa4/ 'hair'). This result revealed cross-linguistic differences.

3.5 Developmental comparison and relation to input forms in Mandarin

The overview of children's two word-points and input speech shows how they have advanced from the restrictions at the onset of word production to the patterns they make as they begin producing more words, and how the ambient language shapes children's speech.

Table 3.23 Prosodic structure in children's production at two word-points and in adult input speech (in %; brackets indicate the sample size)

		CV	CVV	CVC	C_1VC_1V	C_1VVC_1VV CVNCVN	C_1VC_2V	SingleC	longer
4wp (8)	Mean	34	6	NA		55	3	2	
25wp (5)	Mean	25	14	7	25	11	10	11	5
	Sd	8.4	4.2	6.4	8.9	3.9	3.1	6.1	1.7
Input (8)	Mean	15	21	10	4	11	31	8	
	Sd	2.6	1.7	3.0	1.6	1.3	2.9	2.9	

We present the proportions of prosodic structures in actual productions in Mandarin-learning children's 4wp and 25wp and in adult input words. Concerning CV structure, 34% is found in the eight children's 4-word point productions, 25% in the five children's 25-word point productions and 15% in eight caregivers' input speech. The fact that the proportion of CV structures in children's first words is twice as high as in adult input speech but drops in children's later words indicates that the children's forms gradually change in the direction of the input speech as their lexicon grows. The key difference between the first and later words in monosyllables of the Mandarin children is the expansion of the monosyllabic structure to include diphthongs (CVV), which is rare at the 4wp, accounts for more than 10% of all variants overall at the 25wp (Table 3.23) and comes closer to the input proportion.

Turning to disyllables, there are dramatic changes between monophthongal and diphthongal reduplicated structure: CVCV is the most used structure for children at the 4-word point, and the diphthongal pattern is not distinguished from monophthongal due to its low occurrence. At the 25-word point, children produce CVCV only half as much as earlier although still far

more than adult input speech; they use more diphthongal reduplicated words (CVVCVV), matching the proportion of adult input speech.

For variegated words (C_1VC_2V), the low level of consonant variegation seen in Mandarin children's first words (and in target words) is also reflected in their later production. Some children are already beginning to accurately produce words with two different consonants at the later lexical point. Variegated words are mainly avoided by these children or achieved through adaptation to reduplication despite more variegated words being targeted, while adults in our data have a relatively high proportion (one third of all variants) of consonant variegation.

Table 3.24. Consonant manners in children's production at two word-points and in adult input speech (in %; brackets indicate the number)

	Stop	Fricative/Affricate	Nasal	Approximant
4wp (8)	53	8	22	16
25wp (5)	43	23	18	16
Input (8)	33	39	12	16

A preference for fricatives and affricates is apparent in the Mandarin-learning children's first words and is shown at the 25wp for all the children. Thus we see some continuity between earlier and later words. The cross-linguistic differences in frequencies of consonantal manners of articulation identified in input speech (Table 3.24) are apparent in the children's word forms.

In conclusion, Mandarin children's patterns reflect both output constraints and growing systematicity. Mandarin-learning children have been found to acquire diphthongs and fricative/affricate sequences with little difficulty and show considerable developmental advance.

3.6 Discussion and conclusion

The prosodic and segmental challenges that early word production poses to Mandarin children has been investigated in this chapter. First, we focused on the Mandarin children themselves. We analysed the prosodic structure and segmental resources of early words in each child, between the beginning (4-word-point) and end (25-word-point) of their single-word stage. We also analysed the patterns that Mandarin children actively use in relation to the templatic behaviour reported in other languages. Then we made a quantitative comparison with word use by children learning six different languages. We also investigated the input language to see how the adult language shapes children's early words. The cross-linguistic comparison of young children's production of prosodic structures and consonants suggests an interaction between the universal constraints imposed by their immature motor control and memory systems and the language-specific structures and segmental inventories.

Universal constraints and language-specific in prosodic structure

We began the chapter with Mandarin-learning children's first words (4-word-point) and their targets, to see which resources are available to them at the onset of phonological development. We found that the children's first words (4-word-point) are one or two syllables long and mainly take the forms CV and CVCV. Within the routine words input, the first words Mandarin-learning children produced are highly recognizable and matched to their target forms, they may cautiously 'choose' the target forms for which they have a rough vocal match in repertoire. These are broadly similar to those of children learning other languages (Vihman, 2019).

Then we considered the differences in preferences for prosodic structures between the two data points. When Mandarin-learning children's spontaneous word production reaches about 25 word types in a half-hour session, they began to produce more challenging structures, including words with diphthongs, codas and more than one consonant type. However, their criterial use of prosodic structures is still limited to just a few structures (CV, CVV and C_1VC_1V). Mandarin-learning children produce the lowest proportion of words with supraglottal consonant changes compared to other languages. In that sense, their productions are distinguishable from those of children learning other languages.

CV and CVCV seem to serve very generally as the most used structures in Mandarin

children's forms at both 4- and 25-word points. On the one hand, the high proportion of CV and CVCV in Mandarin children's productions may reflect a property of early words and a universal tendency. There are language-general physical constraints, both articulatory and perceptual, on word production in early period and limitations on phonological memory. Based on previous studies, CV syllables are characteristic of children's first words (Stoel-Gammon, 1985). MacNeilage and Davis (2000) and Davis et al. (2002) have promoted the argument of motoric control, connecting children's first words to the evolution of language and claiming that the syllabic structure of language has developed, phylogenetically, from the repetitive movements of the jaw and the tongue during the process of digestion. Reduplicating a syllable (CVCV) is widespread in child phonology and is well illustrated by children learning Jordanian Arabic, German, Japanese, Russian, Swedish and so on (Rose, 2018). Final consonant deletion is another common pattern in the production of children learning many languages (Macken, 1979). Mandarin children here either attempt fewer CVC words or drop the final coda.

On the other hand, the ambient language Mandarin structures also shape Mandarin children's word forms. The syllable structure of Mandarin is much simpler than that of English. CV and CVV monosyllabic structures make up one third of our Mandarin adult input sample. Therefore, it is not surprising that the CV pattern is used most frequently at both lexical points by Mandarin-learning children. The language-specific effects of linguistic exposure also result in more diphthongs being produced in both monosyllables and disyllables by Mandarin-learning children than children learning other languages. Moreover, the infrequent use of the CVC pattern in Mandarin children can be traced to the low proportion of CVC in input speech.

Language-specific effects have also appeared in other aspects. Mandarin children produce the lowest proportion of words with supraglottal consonant changes across the word of all the language groups looked at here. They either avoid words with more than two consonants and restrict themselves to reduplicated words or adapt such words to reduplicated patterns or sometimes produce them accurately for frequent, unavoidable variegated words in everyday life. Certain aspects consistently appear in these word productions. First, the low level of consonant variegation in Mandarin children's production shows it is particularly difficult to retain the adult words which the Mandarin children begin to attempt. Phonological memory is an important factor to recall unfamiliar sound sequences or word forms. Therefore, Mandarin

children increase their vocabulary cautiously through reduplication, within the limits of their ability to retain or represent word forms. According to Singh (2005), reduplication is 'a morphological process that allows speakers to form new words from old words by adjoining the latter in their entirety (complete reduplication)' (p. 263).

Second, there is a low lexical demand for consonantal variegation for Mandarin children at this developmental point. The presence in Mandarin of many vowels per syllable might be sufficient for children to increase their word production while remaining safely within the bounds of a few manageable structures. Meanwhile, it appears to be a characteristic of early words in a language with a great many simple reduplicated words in infant-directed speech. Reduplication in Mandarin is found in many nouns, especially in kinship terms:
e.g. /ma1ma0/ 'mummy' and /pa4pa0/ 'daddy', or in verbs denoting affection:
e.g. /pao4pao0/ 'hug hug' and /tsou3tsou0/ 'go go'. Many of these constitute children's early words. In summary, heavy use of a small number of structures seems to provide sufficient support for Mandarin children at this point.

Third, this could also be a consequence of the parents' feedback in conversation. It is well known that adults accommodate their speech in many ways when addressing young children (Snow, 1972; Cruttenden, 1994; Englund, 2005). Although the proportion of variegated words is higher than that of reduplicated words in our Mandarin input data, many variegated words in Mandarin are reduplicated in infant-directed speech. Loss of contrast is unlikely to affect comprehension, for instance, /t^hu4t^hu0/ for /tu4tsi0/ 'bunny', /pei1pei1/ for /pei1tsi0/ 'cup' and so on.

The discontinuity or regression is also seen in prosodic structures. At the 4-word point, children focus on the words available for their repertoire. CV and reduplication (CVCV) are the main prosodic patterns for Mandarin children at this developmental point and these tend to perfectly match the target patterns. When moving to the 25-word point, children attempt more challenging prosodic patterns, such as CVV and variegation. Some patterns previously characteristic of some children are no longer available. Some of the children who used the most reduplication (for example, Yiyi and Keke) at the 4-word point no longer do so as their main pattern at the 25-word point.

Finally, children's word forms become less accurate but more in line with their current phonological patterning.

Universal constraints and language-specific in consonants

Children's consonant acquisition is another important feature of their overall phonological development. Universal tendencies have been found in the consonant inventory among Mandarin-learning children and cross-linguistically. Stops and nasals are the most common patterns for all Mandarin-learning children at the 4-word point. At the 25-word point, stops are still the most used consonant for Mandarin-learning children, just as is the case for children learning other languages; however, fricatives and affricates (23%) are also produced frequently.

The acquisition of fricatives and affricates shows the influence of the ambient language. Late acquisition of fricatives and affricates has been documented in many languages (Locke, 1983). The fricative-vowel sequence is more demanding than other sequences, in that the jaw must lower in anticipation of the vowel while the tongue remains high. These complex dissociated movements are a challenge for children. But these typically late-acquired sounds emerge early in Mandarin children's production. Some Mandarin-learning children produce fricatives and affricates in their first words (4-word point) in relevant every-day contexts (e.g. /tɕiɛ3tɕiɛ0/ 'sister'; /ɕiɛ4ɕiɛ0/ 'thanks'). Mandarin-learning children produce the most fricative and affricate sounds among the language groups we compared at the 25-word-point (Table 3.19). Beyond that, our data also shows that these children do not always replace fricatives and affricates with sounds that are considered easier to articulate. One child (Shi) uses $C_{tɕ}V_{FRONT}$ as an active template. This recalls Zhu's (2002) finding that 'in the youngest age group (1;6-2;0), 90% children were able to articulate the affricates [tɕ] and [tɕ^{h}] once and several times' (p.66). This indicates that articulatory constraints interact with the ambient language in Mandarin children's segmental acquisition. The differences between Mandarin and the other three languages (English, French and Swedish) in the frequency of occurrence of fricatives and affricates in the input indicates that this adult difference leads to a difference in children's production by the end of the single word period (25-word point). Mandarin-learning children increasingly develop their production of fricatives/affricates between the two word-points. This advance illustrates both ambient language effects and developmental continuity.

Individual differences

There are individual differences in the words of children learning Mandarin in the same community. The wide range and large standard deviations in proportions of prosodic structure at the end of single word period (25wp) demonstrate high individual variability within groups. Regarding the two most used prosodic patterns (Table 4.3.3h), the children ranged in CV syllable production from 16% (Xinyu) to 38% (Shi). The proportions of CVCV structure ranged from 11% (Yiyi) to 33% (Shi). The active templates are different in each Mandarin child. This suggests that children's actual productions might be more affected by their individual phonological preferences, for example, templates (Vihman & Croft, 2007), than by the adult targets.

In conclusion, Mandarin-learning children start by producing, in relevant routine contexts, simple words which are broadly similar to their adult pattern and to children's patterns in other languages, then later begin to overcome the challenges by selecting and adapting target words to fit their existing patterns. Frequency of occurrence in the input language plays a key role in shaping Mandarin-learning children's patterns at this developmental point. The child's segmental inventory is a factor which interacts with prosodic structure in template use.

4. THE CHALLENGES OF LEARNING TONE IN EARLY WORDS

4.1 Introduction

Besides the fact that prosodic structure and segment are essential aspects of Mandarin phonological system, as of any European language, the most fascinating aspect is lexical tone. The focus of this chapter is on the development and change of lexical tones over time in the word learning period of Mandarin-speaking children. Based on the description of Mandarin tonal phonology in Chapter 1, this chapter is primarily intended to systematically describe the nature of tone acquisition, including lexical tones and weak syllables (neutral tones) in the 8 monolingual Mandarin-learning children at two single-word developmental points (4-word point and 25-word point). Taking both child actual tonal patterns (occurrence) and target adult tones (accuracy) into account, this chapter illustrates the effects of universal constraints and the ambient language on early tone production, tone selection and the pathway to mature tone production in the Mandarin-learning child and discusses possible reasons why tone may develop differently than vowels and consonants.

The chapter begins with a review of previous studies on lexical tone acquisition from perceptual and production perspectives, and on the development of neutral tone (or weak syllable). It then shows the implications and limitations of these results and moves on to present the current project's analyses and findings.

4.1.1 Previous Studies of Lexical Tones

Lexical tones are used as phonological categories in addition to vowels and consonants to draw meaningful distinctions between words. Thus tones constitute an important factor in lexical retrieval for adults (Cutler & Chen, 1997) as well as children. Lexical tones are found in most of the languages native to Sub-Saharan Africa, as well as in East and Southeast Asia (Yip, 2002). However, research on phonological development has largely focused on European languages. Tone-language-learning in children has received much less attention (Singh & Fu, 2016). The development of lexical tone has remained relatively unexplored to date.

In studies of vowel and consonant perception, infants are documented as starting with the ability to discriminate a variety of contrasts as "universal listeners", then shifting to being attuned to native phonological contrasts in the first year of life: This shift takes place by the age of six to nine months for vowels and by nine months for consonants (Werker & Tees, 1984, Polka & Werker, 1994; Kuhl et al., 1992; Kuhl, 2011).There have been relatively few studies investigating when infants attune to tone.

The first study was conducted by Harrison (2000), who compared tone-language-learning (Yoruba) infants with non-tone-language-learning (English) infants of 6 to 8 months on responses to Yoruba tone contrasts. Then found that only the Yoruba infants achieved any level of statistical significance in their response to pitch differences in isolated syllables. This study revealed that early tone discrimination is modified by native-tone language experience. By age 6-8 months the infants in Harrison's study had already attuned to native categories.

A somewhat different and more nuanced result was presented by Yeung, Chen and Werker (2013). They investigated developmental changes in tone perception by comparing the discrimination of Cantonese tones in infants learning tone languages (Cantonese or Mandarin) and non-tone language (English) at 4 and 9 months. English infants demonstrated a decline in sensitivity, in that they could discriminate Cantonese tone contrasts at 4 months but no longer at 9 months. By contrast, Cantonese- and Mandarin-learning infants could discriminate Cantonese tones over the entire period. This study found language-specific effects on infants' tone perception as early as 4 months of age. English, Cantonese, and Mandarin infants each demonstrated different discrimination abilities in accordance with the properties of their native language at this stage. For example, 4-month-old English infants looked longer at Cantonese Tone 33 trials, compared to Mandarin infants, whose looking time to Cantonese Tone 25 trials were longer; whereas Cantonese infants showed an equivalent looking time to both tone trials.

Although tone categories seem to emerge early in development, that leaves a question as to the sensitivity of infants raised with one tone system to tones from a different language system. Mattock and Burnham (2006) using Thai tone contrasts tested Chinese- (Mandarin or Cantonese) and English-learning infants' sensitivities at 6 and 9 months. They reported that English infants exhibited a decreasing sensitivity, while Chinese infants showed a sustained sensitivity to Thai tones between 6 and 9 months. Meanwhile all groups showed similar

sensitivities to musical tone. This suggest tone attunement may generalize across tone inventory.

Within native tones, is it that infants demonstrate similar perceptual sensitivity for different tone contrasts? Could the acoustical properties of tone contrasts affect native lexical-tone perception? Moreover, directional asymmetries are reported in vowel discrimination: 6 to 8 and 10 to 12-month-old infants showed directional asymmetries for native vowel contrasts (Polka & Bohn, 2003). Therefore, Tsao (2008) tested 10- to 12-month old Mandarin-learning infants on four Mandarin tone contrasts. The tone pairs to be discriminated varied in acoustic salience. Although Mandarin Chinese infants were able to discriminate Mandarin tones, they showed differences in performance based on the tone pair tested: the most acoustically distinct contrast (i.e., Tone 1 vs. 3) is easier to discriminate than the other two contrasts (Tone 2 vs. 3 and Tone 2 vs. 4). There is an effect of acoustical similarity on lexical-tone discrimination in Mandarin-learning infants. In addition, there is a directional asymmetry of lexical tone perception: Tone 1 → Tone 3 is the easier direction for infants to discriminate than Tone 3 → Tone 1.

Lexical tones change their phonetic property in connected speech. The latter notion was supported by Shi (2010), a study testing Chinese infants on native tonal contrasts in various tonal contexts. The study showed that Mandarin-learning infants at 8 to 11 months, but not at 4 to 6 months, can discriminate native tone contrasts in those variable contexts, suggesting that perceptual abilities of tone for native infants do not emerge until at least 8 months of age.

Since every language incorporates pitch contrasts at the level of intonation, the way non-tone-language learners perceive lexical tones has also received attention. There are a few studies of non-tone-language learners perceiving lexical tones. Mattock et al. (2008) investigated 4-, 6-, and 9-month-old English- and French-learning infants' sensitivity to Thai tone contrasts. The authors found that both French- and English-learning infants discriminated those tones at 4 and 6 months, but not at 9 months of age.

However, a recent study (Liu & Kager, 2014) claimed that the decline in tone discrimination in non-tone language learners is transient and reversible. Liu and Kager (2014) tested Dutch infants in several age groups between 5 and 18 months. Infants' sensitivity to a Mandarin tone contrast (Tones 55 and 51) demonstrated a U-shaped pattern: infants showed relatively

strong discrimination prior to 8 months and after 12 months but the weakest discrimination between 8 and 12 months. Once stimuli become harder to contrast, only the youngest (5–6 months) and oldest (17–18 months) infants succeeded in discrimination.
In addition to developing perceptual abilities, tone language learners need to produce lexical tones accurately. Prior studies of tone production are mainly based on only a few children. Although researchers examined tone errors and the age and order of acquisition of individual tones, they do not distinguish between what children produced regardless of the target (occurrences) and what they produce given a particular target (accuracy). There are disagreements concerning the order of acquisition of the tones.

Clumeck (1977) documented the production of lexical tones in one infant exposed to Mandarin in the second year of life. He found that his subject used rising pitch for all disyllables when he began to use words at about the age of 1;10. Although the high level and falling tones appeared about one month later, they were sometimes substituted by rising tone. Clumeck (1980) observed another two children later at 2 and 3 years old, respectively. He found both children could produce all Mandarin tones over the observational period.

Likewise, Li (1977) and Hsu (1987) reported that their subjects easily acquired high level, rising, and falling tones, but had considerable difficulty with falling–rising tones. In contrast, Li and Thompson (1977) investigated 17 Mandarin learners' tone production at ages 1.5 to 3 years using a picture-naming task. They found children primarily correctly produced high level and falling tones; while they confused rising tone and fall-rise tones for a relatively long time, even when they were already in the multiword production stage. When they moved to the sentence stage, they could produce the correct tone with each word. In Wong's studies (Wong, 2013; Wong et al., 2005), the acquisition order from earliest to latest is: falling tone, level tone, rising tone, and finally fall-rise tone. Within this order, there is still a sharp decrease for fall-rise tone.

A recent study of tone acquisition using a large group sample was conducted by Zhu and her colleague (Zhu & Dodd, 2000; Zhu, 2002). A cohort of 129 monolingual Mandarin-learning children were observed between the ages of 1.5 and 4.5 years in a cross-sectional design. The results demonstrated limited tone errors, even in the youngest group of children. The only two tonal errors were cases in which rising tone within some specific segments were replaced by falling tone and falling-rising tone were replaced by rising tone. Zhu and her colleague

concluded that tone is the first type of phonological category to be acquired.

However, we cannot conclude that children have sophisticated knowledge of tone in early life. A collection of subsequent studies of tone production using adult judgment and acoustic analysis showed that lexical tones emerge relatively early but take several years to mature and reach adult-like accuracy. Wong, Schwartz, and Jenkins (2005) analyzed 3-years-old children's speech using a picture-naming task. The results showed that the children could accurately recognise most tones in monosyllabic words, but produced qualitatively different and significantly fewer accurate target tones than did adults. A follow-up analysis with older children demonstrated that tone is still not completely acquired by 4–5 years of age (Wong, 2013). Wong and her colleagues' findings contrast with early studies which demonstrated accurate tone production within first two or three years of children's life.

4.1.2 Acquisition of rhythm in non-tone languages and Mandarin

Producing adult-like rhythm is another difficult task for the prelinguistic infant. Rhythm is seen as a regulator of motor behavior in speech production and a link between biological constraints and linguistic structure (Kent, Mitchell & Sancier,1991). The terms 'iambic', 'trochaic' and 'spondaic' are used to describe weak-strong, strong-weak and even-stress patterns in adult languages, respectively. The phenomenon of final syllable lengthening seems to be a natural phonetic feature and general tendency in infants' early speech. Therefore, in trochaic words of a language, children need to learn to shorten their final syllables. Studies have documented limited control of timing in infants' early years of life. Adult-like duration is acquired only at later developmental level.

Laufer (1980) reported final syllable lengthening in four English infants' (5-12 months) production. Robb and Tyler (1995) analyzed vocalizations in seven English-learning children from 8–26 months. They found a decrease in word duration with age, suggesting a maturational effect on word duration over time. Kehoe, Stoel-Gammon, and Buder (1995) also found that 18-month-olds produced more adult-like pattern in the use of stress in trochaic words in terms of f0, intensity, and duration.

Cross-linguistic studies which contrast the 4- and 25-word point also provide evidence that adult-like duration is only achieved at the latter developmental point. Vihman and DePaolis

(1998) found the vowel durations of French and English infants were not systematic or adult-like at the 4-word point, while infants produced more adult-like patterns by the 25-word point. Vihman, Nakai, and DePaolis (2006) tested the duration of VCV disyllables in English-, French- and Welsh-acquiring children and found most children in all three groups lengthened the final syllable but did not show ambient language effects at the 4-word point, while by the 25-word point all three groups began to produce adult-like rhythmic patterns. In the study of investigating geminates versus singleton stops, Finnish infants were found much closer to the adult models in their production of the 25-word point than that of the 4-word point.

All these studies suggest that there is limited control of prosody in the pre-linguistic or first word period, although early ambient language effects are seen in infants' pre-linguistic vocalizations. Final-syllable lengthening could be due to physiological constraints. The control of duration may involve several stages of learning over the course of development with increasing motor control. As the infant approaches the second half of the second year of life, the prosody of vocal productions become increasingly adult-like. In other words, after the acquisition of a sufficient number of words, infants begin to integrate adult-like prosody into their vocalizations. The changes in production could be due to a loosening of physiological constraints, or to growing awareness of prosodic or segmental contrasts.

De-stressing is reflected in the process of Mandarin learning. Note that the weak syllables are notoriously difficult to handle for Mandarin learning children. Sometimes children produce the words which should be unstressed as stressed, and the words which should be stressed as unstressed.

The neutral tone plays an important role in the speech rhythm of Mandarin. Similar to children acquiring non-tone languages, Mandarin children are also found to only slowly begin destressing the weak syllables (so-called 'neutral tone') and do not master this until around age 4-6 years. Li and Thompson (1977) observed that the neutral tone of weak syllables was often replaced with a full lexical tone in children's speech during the process of acquisition, e.g., using /tsi3/ for /tsi0/. Su (1985) suggested that children might use different strategies in the substitution of full tones for weak syllables, depending on different contexts. The weak syllables occurring in utterance-final particles, classifiers, particle 'de', and nouns ending with 'tsi' were likely to be replaced with low falling tones, while the weak syllables in

disyllabic words were likely to change into full tones. All these types of weak syllables appeared in previous studies based on recordings of older children, while only reduplication can be observed in the current study in the children's production at the 4wp. Younger children in Hua (2001) more often replaced weak syllables by citation tones, that is, they produced more even-timed syllables than older children, such as /t^hou2 fa4/ for /t^hou2 fa0/ 'hair'. The acquisition of weak syllables in Mandarin reflects interactions between lexical and phonological development. Children in Hua (2001)'s study made progress both in the number of usage and in accuracy and started to achieve about one-third accuracy from age 4.

4.1.3 The current study

First, the early studies in Mandarin were limited by their focus on a single child each in diary studies or small sample sizes and single observation in cross-sectional studies. Thus there is not enough data recording children from the same language environment. More important, early studies used only accuracy of production to measure tone acquisition. They gave too much attention to Mandarin-learning children's phonetic competence, that is, their achievement of a correct match between the child's tone production and adult target. Little consideration was given to the child's actual production of tone. This did not allow them to detect the systematic change in the use of tone in children's early productions. Therefore, occurrence, accuracy and variability of tone will all be explored in this study.

Second, the gap between linguistic structure as classically described and the actual input signal as experienced by infants is highlighted by the fact that the prosodic cues in infant-directed speech are far more variable than those in isolated disyllables elicited from adults (Vihman et al., 2006). It has been shown that, in English, infant-directed speech is much more variable than adult-directed speech, which provides a poor teaching signal (Vihman et al., 2006). How about other languages, such as Mandarin? This chapter will examine whether the patterns found in infant output are due to a lack of consistency in the input, the inability of the infant to control the motor system, or some combination of the two.

The chapter aims to 1) examine the challenges of tonal pattern that Mandarin children face in the period of early word production by analysing children's tonal patterns at two lexical points; 2) explore how universal constraints, ambient language and adult input speech affect Mandarin-learning children's tone production by looking at the changes in tone production

with development over time – that is, which patterns emerge and which ones fade.

4.2 Methods

The York Mandarin-learning children who were studied in Chapter 3 were assessed regarding the use of lexical tone. The use of tone in each child's spontaneous word production was transcribed and counted for the relevant session. Two lexical points (the 4 word-point and 25 word-point) were also applied here. Onomatopoeia and exclamations were all counted as words in Chapter 3 (see Vihman 1996), but not here: some exclamations are not included (e.g. /ai1ya0/ 'uh-oh') due to different tonal input, family by family. Only tones in mono- and disyllabic words were analysed. Monosyllabic or disyllabic words constitute the majority (94.14%) of word types in the Mandarin CHILDES database (Wang et al., 2010). Target tone was defined as the adult pronunciation of the tone in the word the child was attempting to produce.

Since the data were based on naturalistic speech production, whether a child has a chance to use a target tone varies from child to child and session to session. Therefore, the non-occurrence of a tone in a child's production at any point may not mean that the child could not produce it. This should be taken into consideration when interpreting the results.

Our analyses focus on two perspectives: child-centered (i.e. tones produced) and adult-centered (i.e. tones attempted and accuracy). We distinguish occurrence of each tone from accuracy in tone use in order to highlight the possible changes between the first words and later words, and to identify emergent system.

4.3 Data Analysis and results

4.3.1 Tonal patterns in Mandarin-acquiring children at the 4-word point

Occurrence

We first look at the usage of lexical tones in the first word production period (the 4-word-point) of 8 Mandarin-learning children (Table 4.1). We will start from tones in monosyllabic words (the top rows in the table). In monosyllables, Tone 1 (42%) and Tone 4 (42%) are the most common patterns used by these Mandarin children. All the children produce words with Tone 1; almost all children produce words with Tone 4. There are fewer uses of Tone 2

(15%), which occurs in three children's monosyllabic vocalizations at this developmental point. One child (Didi) uses Tone 2 but does not produce Tone 4 in this session. None of the children produces Tone 3. Note here we only discuss tones children actually produced, disregarding the target tones.

Table 4.1 Occurrence of tones in mono- and disyllabic words and number of matches. Type/Tokens in total (types with at least one token matching for tone; **bold face** means complete match).

	Keke	Xinyu	Yifan	Yiyi	Zuozuo	Didi	Shi	Weilun	Total	
Mono	6	6	2	5	7	3	2	2	No.	%
Tone1	2/2 (1)	**2/3 (2)**	1/4	3/7(1)	3/4 (1)	1/2	**1/4 (1)**	1/1	14	42%
Tone2		**1/1 (1)**			**2/3 (2)**	2/5 (1)			5	15%
Tone3										
Tone4	4/5 (2)	3/7	**1/1(1)**	2/7(1)	**2/2 (2)**		**1/3 (1)**	1/2	14	42%
Di	9	6	4	8	4	11	5	8	No.	%
T1-T0	2/2 (1)	2/3		**1/1(1)**	**1/8 (1)**	2/7 (1)	1/2 (1)	2/3(1)	11	20%
T1-T1			**1/2(1)**	2/5 (1)		**1/1 (1)**	**2/2 (1)**	1/1	7	13%
T1-T2					1/1				1	2%
T1-T4	2/5	2/4	2/4	3/6	1/1	1/1		3/4	14	25%
T2-T2					1/1				1	
T2-T4	1/1			1/2			1/1		3	

T4-T0	**1/1(1)**			**1/1 (1)**		**3/9 (3)**	**1/2 (1)**		6	11%
T4-T1	1/1			1/1		2/2		1/2	5	
T4-T4	2/2	2/3	1/1			2/3		1/2	8	14%

Now we move to tones in disyllabic words (the lower part of the table). As seen in Table 4.1 and 4.2, Tone1-Tone4 and Tone1-Tone0 are the most two common types, in both total occurrence (25% and 20%, respectively) and the number of children who produce them (7 for both tones). The next most common patterns are Tone4-Tone4 (14%) and Tone1-Tone1 (13%); over half of the children produce words with these patterns. Tone4-Tone0 (11%) is the last most used pattern, produced by half of the children. Other combinations like Tone1-Tone2, Tone2-Tone2 or Tone2-Tone4 are only produced a few times by a few children or by onechild, different for each pattern.

Table 4.2. The high-use tone combinations in disyllable at 4wp (mean occurrence)

Tone combinations	Used by N children
Tone1-Tone4 (25%)	7
Tone1-Tone0 (20%)	7
Tone4-Tone4 (14%)	5
Tone1-Tone1 (13%)	5
Tone4-Tone0 (11%)	4

Accuracy

Regarding matches, the most accurately produced tone in monosyllable is Tone 2 (80%). There are only a few occurrences of Tone 2, but most of them are used to realize Tone 2. The proportions of matches for Tone 1 and Tone 4 are 43% and 50%, respectively: two out of the seven children who produce Tone 4 mismatch all, and three out of eight children mismatch all their Tone 1 productions. We can see that only half of the uses of Tone 4 or close to half

of Tone 1 are accurate; the remaining uses are substitutions for other target tones. In other words, Tones 1 and 4 show considerable overuse. As regards Tone 2, there are three words with Tone 2 that are used by the three children: /we2/ 'hello (on the phone)', /mau2/ 'peekaboo' and /na2/ 'take', all of which are social words used in daily routines. That routine use might help explain why Tone 2, while less used, is the most accurate tone at this point. These Mandarin children exhibit three general tone substitution patterns at this early developmental level. First, Tone 2 (rising tone) was substituted by either Tone 1 (50%) or Tone 4 (50%), e.g., [kɛ1] for /kei2(ni0)/ 'there you go'. Tone 3 (falling-rising tone) is produced as Tone 1 (25%) or Tone 4 (75%). Finally, Tones 1 and Tone 4 substitute for one another.

For disyllables (see Tables 4.1 and 4.3), all productions of Tone4-Tone0, one of the most-produced tones, are a match, for all four children who produce it. It turns out that only two Tone4-Tone0 words are produced: the most used ones are /pa4pa0/ 'daddy' and for one child, also the sister's name /na4na0/ 'Nana'. While Tone1-Tone4 is the most common combination, none of its productions is a match to the target. All the Tone1-Tone4 patterns are produced as substitutes for other target patterns. The proportion of matches of another most common combination Tone1-Tone0 is 55%. As for the two next most-common tonal combinations, none of the Tone4-Tone4 productions match the target, and for Tone1-Tone1 57% are a match. There are no Tone4-Tone4 target words; all the Tone4-Tone4 productions are used to substitute for other tone patterns.

Table 4.3 Summary of tone match in mono- and disyllabic words at 4wp (only most frequent patterns in disyllable are listed here).

Tones in Mono	Matches	Tones in Di	Matches
Tone 2	80%	Tone4-Tone0	100%
Tone 4	50%	Tone1-Tone1	57%
Tone 1	43%	Tone1-Tone0	55%
Tone 3	N/A	Tone1-Tone4	N/A
		Tone4-Tone4	N/A

It seems difficult for the children to produce accurate tonal patterns in disyllabic words (only 29% match on average). Looking more closely at the words produced, the most frequent patterns (in proportions) are T1-T4 and T1-T0. As reviewed previously, neutral tone has different phonetic realisations, based on the preceding tone. It has a rising contour only after Tone 3, but a falling contour after Tone1, Tone2 and Tone4. That is, Tone1-Tone0 means level contour followed by a falling. Therefore, along with T1-T4 (high level and falling), we can conclude that the favourite patterns produced by these Mandarin children are level tones with drops. The difference between T1-T0 and T1-T4 is the prosodic pattern: one is strong-weak (trochaic) and the other is relatively even (spondaic). For the level tones-with-drop pattern, these Mandarin learners apply the trochee (T1-T0) and spondee (T1-T4) patterns a similar amount, but all T1-T4 are inaccurate use or overused. Except for T4-T0 (the second most common pattern) and a single instance of T2-T0, we find that they show a range of different 'spondaic' patterns: T2-T2, T2-T4 T4-T1 and T4-T4, all with no matching uses at all. These Mandarin learners seem to assign equal stress to each syllable in the disyllabic word rather than distinguishing between strong (citation tone) and weak stress (neutral tone) at the 4-word point.

4.3.2 Tonal patterns in Mandarin-acquiring children at 25-word point

We now go on to look at tone use when the remaining five children are producing more than 25 words in a half-hour recorded session. As is apparent from Table 4, children explore more tones in terms of both types and tokens at the 25-word point. Tone 1 (36%) and Tone 4 (40%) are still the most popular monosyllabic patterns in these children, and all children use these two tones at this point. Tone 2 is the next most used (21%) by all five children now, especially Yiyi who produces a large number. Tone 3 is emerging but is produced only three times, by three children.

The disyllabic tones that the children produce at the 25-word point differ by child, as seen in Table 4.4. The children are trying to produce more tonal combinations in disyllables and add complexity. Tonal patterns used by at least four of the five children are listed in Table 4.5. T1-T0 (18%) and T1-T4 (15%) continue to occur as the favorite patterns and are used by all five children. As for the other three high-use patterns at 4wp, T1-T1 (14%) and T4-T0 (12%) are still used commonly by all or almost all children; T4-T4 has decreased in use relative to the 4wp. Citation tone followed by neutral tone like T2-T0, which is a challenging pattern

requiring more motoric skills, is produced by all five children. We did not see this pattern commonly in the early word point. Tone 3 emerges in disyllabic word T3-T4 as well and is used by all the children.

Table 4.4 Tones used by 5 Mandarin children at 25-word point and number of matches. Type/Tokens in total (types with at least one token matching for tone; bold face means complete match).

	Keke	Xinyu	Yiyi	Didi	Shi	Total	
Mono	18	16	26	20	24	104	
Tone1	4/5 (2)	6/7 (3)	9/17 (8)	9/21 (6)	7/21 (6)	35	36%
Tone2	3/4 (1)	1/2 (1)	12/28 (5)	2/3 (1)	2/2 (2)	21	21%
Tone3	1/1 (1)		1/1 (1)		1/1 (1)	3	3%
Tone4	9/11(7)	6/14 (2)	4/10 (2)	5/12 (2)	15/45 (10)	39	40%
Di	30	42	15	35	40	162	
T1-T0	7/21 (3)	10/34 (3)	4/4 (1)	3/3 (2)	5/12 (3)	29	18%
T1-T1	3/3	7/11 (2)	1/1	3/5 (1)	8/8 (2)	22	14%
T1-T2	**1/1 (1)**	2/2 (1)	1/1	**1/3 (1)**		5	
T1-T4	6/8	5/19	3/3 (1)	3/5 (1)	8/16	25	15%
T2-T0	2/5	2/2	2/2 (1)	**1/2 (1)**	2/5 (1)	9	6%
T2-T1		1/1				1	
T2-T2	2/3			1/2	1/1	4	
T2-T3	**1/2 (1)**			**1/1 (1)**		2	
T2-T4		**1/1 (1)**	2/2	2/2 (1)	4/4 (1)	9	6%
T3-T0				1/3 (1)		1	
T3-T1	1/1	1/1		5/20 (3)	1/1	8	

T3-T2			1/1			1	
T3-T4	**1/1 (1)**	**1/1 (1)**	1/1	5/9 (1)	2/2 (1)	10	6%
T4-T0	3/7 (3)	7/15 (5)		6/10 (3)	3/4 (1)	19	12%
T4-T1	1/1				1/3	2	
T4-T2	1/2	2/2				3	
T4-T4	1/1	2/4	1/1	3/3 (2)	4/5 (1)	11	7%
T4-T3		1/1				1	

Table 4.5 The high-use tone combinations in disyllables at 25wp (mean occurrence)

Disyllabic Tones produced	N children
Tone1-Tone0 (18%)	all 5
Tone1-Tone4 (15%)	all 5
Tone1-Tone1 (14%)	all 5
Tone4-Tone0 (12%)	4
Tone4-Tone4 (7%)	all 5
Tone2-Tone0 (6%)	all 5
Tone3-Tone4 (6%)	all 5
Tone2-Tone4 (6%)	4

Consider now the accuracy of tones used in monosyllables. Surprisingly, Tone 3 is the most accurately used tone at this point. Although there are only three items with Tone 3 produced by three children each, they are perfectly matched to target (100%). More Tone 1 (70%) and Tone 4 (59%) are used to resemble the target patterns. With the increasing occurrence of Tone 2, the accuracy falls (48%). With still relative few productions, only a few Tone 2 uses

are matches to target, while most are substitutions for Tone 3, for instance [ma2] for /ma3/ ‘horse’, [wɤ2] for /wɤ3/ ‘me’, and [ʂiu2] for /ʂou3/ ‘hand’, etc.

Tone errors are analysed at this developmental level. We can see three strategies of tone substitutions: Tone 2 was misused as either Tone 1 (43%) or Tone 4 (57%); Tone 3 is substituted by Tone 2 (75%) or Tone 4 (25%). Finally, Tones 1 and 4 substitute for one another, but the proportion decreases. At this point, Tone 1 no longer substitutes for Tone 3.

Table 4.6 also summarises the proportion of matches in disyllables. Besides Tone4-Tone0 and Tone1-Tone0, which are still used with relatively high accuracy, the emerging patterns which involve Tone 2 or Tone 3, although not very frequent, are highly accurate.

Table 4.6 Matches of tone use in mono- and disyllabic words at 25wp (only high match in disyllable are listed here).

Tones in Mono	Matches	Tones in Di	Matches
Tone 3	100%	Tone2-Tone3	100% (2/2)
		Tone3-Tone0	100% (1/1)
		Tone4-Tone0	63% (12/19)
Tone 1	70%	Tone1-Tone2	60% (3/5)
		Tone1-Tone0	41% (12/29)
Tone 4	59%	Tone3-Ton4	40% (4/10)
		Tone3-Tone1	38% (3/8)
Tone 2	48%	Tone2-Tone0	33% (3/9)
		Tone2-Tone4	33% (3/9)

4.3.3 Developmental change and the emergence of a system

The gender of the infants, the age on attaining the different word points, and the number of word tokens were analysed. The mean age of reaching the criteria for the 4-word point was

12.2 months, and that for the 25-word point was 17.4 months. The number of tone types in monosyllabic and disyllabic words, and the mean percentages were calculated at the two word-points in order to examine the distribution of tones. An examination was made as to how the distributions of the tones changed with increasing lexical growth.

Occurrence

Having examined the tone pattern variants produced at each word point we now summarise briefly the tone production of the children at the two word-points (table 4.7). Both children's productions (occurrence) and matches to the target tones (accuracy) have been analysed in order to examining the relationship between lexical and phonological development. We also see some continuity between earlier and later words: the child who is found using Tone 2 at 4pw session definitely produces Tone 2 at 25wp session. There was a decrease in the proportion of both Tone 1 and Tone 4 in monosyllables as children's lexicon grew larger at 25wp, especially for Tone 1. This is because more Tone 2 is produced (from 15% to 21%). This is also because the emergence of Tone3 at the 25-word point: none of the children produced words with Tone3 at the 4-word point sessions, but 2% of the produced words were Tone3 at the 25-word point.

The percentage of tone patterns in disyllabic words showed a change with increased lexical growth (Table 4.7). Two trajectories are observed for the most popular tones of the 4-word point: two trochaic patterns (here meaning any citation tone followed by neutral tone), namely, T1-T0 and T4-T0, remain at roughly the same proportion; while another two favorited spondaic patterns, namely T1-T4 and T4-T4, show a decrease. The less frequent tonal patterns of the 4-word point are increasingly produced. Noticeably, T2-T0 and T3-T4 were never produced at the 4-word point but come up to 6% at the 25wp. Analysis of longitudinal data identified an important change between the two word-points.

Table 4.7. The distribution of 4 tones in monosyllables and frequent-use tones in disyllables at two word points (ordered by the occurrence of 4-word point)

	4-word point		25-word point	
Tone1	14	42%	35	36%
Tone4	14	42%	39	40%

Tone2	5	15%	21	21%
Tone3	N/A		3	3%
T1-T4	14	25%	25	15%
T1-T0	11	20%	29	18%
T4-T0	6	11%	19	12%
T1-T1	7	13%	22	14%
T4-T4	8	14%	11	7%
T2-T4	3	5%	9	6%
T2-T0	N/A		9	6%
T3-T4	N/A		10	6%

Accuracy

By the 25wp there is an increase in the accuracy of Tone 1 and Tone 4 compared to what was found at the 4wp (see Table 4.8). However, less accuracy was observed for Tone 2. Tone 2 was used accurately, for Tone 2 words, when it emerged at the 4-word point; with increasing Tone 2 use as a result of lexical growth, as well as continued vocal practice, there is a regression in accuracy. Meanwhile, Tone 3 words emerged at the 25-word point, and they are all matched to the target tone.

Only three tonal patterns in disyllabic words used have some matches at the 4-word point: T1-T1, T4-T0 and T1-T0 (from highest to lowest matches). Moving to the 25-word point, children become more familiar with these three patterns through the vocal practice (slightly increase or same proportion in production), but they now use them much less accurately. The frequently produced but inaccurate patterns at the 4wp, for instance, T4-T4 and T1-T4, are produced much less at 25wp. However, when they are produced, their accuracy increases, although it is still low for T1-T4. As for the new patterns involving Tone 2 and Tone 3, two of them, T2-T3 (100%) and T3-T0 (100%), are used accurately each time, and there is a relatively good match for other new patterns.

Table 4.8 The accuracy of tones in monosyllables and disyllables at two word points (ordered by the accuracy of 25-word point and only matched categories in disyllable are listed; - =not produced at all, 0= produced but does not match target)

	4-word point	25-word point
Tone3	-	100%
Tone1	43%	70%
Tone4	50%	59%
Tone2	80%	48%
T2-T3	-	100%
T3-T0	-	100%
T4-T0	100%	63%
T1-T2	-	60%
T1-T0	50%	41%
T3-T4	-	40%
T3-T1	-	38%
T2-T0	0	33%
T2-T4	0	33%
T4-T4	0	27%
T1-T1	57%	23%
T1-T4	0	8%

Realizations of target tones

The accuracy we described earlier is child-centred; that is, we indicated whether or not the tonal pattern that the child produced resembled the correct target tone. Now we turn to an adult-centred perspective. We consider the tones that the children attempt and whether these attempts match the targets or are substituted by other tones. Table 4.9 presents the results of

this analysis: the frequency of the target and substitution patterns for the tone attempted. This also reveals an important developmental change between the two word-points. Over half of the attempts at Tone 2 are realised as Tone 2 at the 4wp; the rest are replaced by Tone 1 or Tone 4, with Tone 1 being the more frequent realization. The same strategies are applied at the 25wp, but the accuracy of realization of Tone 2 declines. At the 4wp, Tone 3 is not realized successfully at all and is either replaced by Tone 1 or Tone 4. However, interestingly, Tone 1 is not used as a substitute for Tone 3 at the 25wp; Tone 3 is replaced by Tone 2 and Tone 4, which have acoustic properties related that of Tone 3 in tone sandhi and in running speech.

Table 4.9 Realisation of attempted tones at two words points

Target	Realisation	4-word point	25-word point
Tone 2	Tone 2	57%	42%
	Tone 1	29%	25%
	Tone 4	14%	33%
Tone 3	Tone 3	N/A	11%
	Tone 1	25%	N/A
	Tone 2	N/A	67%
	Tone 4	75%	22%

4.4 Lexical tones in input speech

What leads infants to settle on particular patterns, such as words in Tone 1 and Tone 4, which are dominant here? The phonology of target words has been shown to affect early vocabulary acquisition. Target words could be expected to reflect some aspects of the children's production capacities, which might be closer to the children's word shapes than to a sample of the maternal speech. If we consider the distribution of tones in target words (see Table 4.10), Tone 2 and Tone 3 words are attempted almost as much as Tone 1 and Tone 4 words. A version of the Mandarin CDI (Hao, Shu, Xing and Li, 2008) is also analysed (only Part 1 that

covers infants aged 12 to 16 months which equivalent to the age range of our participants is analysed). The phonological analysis shows that the four tones in monosyllabic words are relatively equally distributed over the words in the CDI (which are a mix of words children may attempt and ones they may comprehend from the input): Tone 1: 27, Tone 2: 24, Tone 3: 25 and Tone 4: 27. However, the tones produced in monosyllabic wordsin our study are largely restricted to Tone 1 and Tone 4, as seen in Table 1.

Table 4.10 Attempted tones in monosyllabic target words at two lexical points and tones in CDI

	Tone 1	Tone 2	Tone 3	Tone 4
4wp target	6	4	4	6
25wp target	26	11	14	24
CDI (Hao et al.,2008)	27	24	25	27

Is it possible that the speech directed to the infant plays a role in the perceptual salience of those tones? Hart (1991) found that the first words used by 45 English-learning children tend to be those most frequent in parent use. In adults, word frequency affects processing in both perception and production: high-frequency words are associated with faster word recognition and are produced more quickly and more accurately (see Ellis, 2002). Therefore, the speech adults address to their children needs to be analysed. To provide an estimate of tonal types in the ambient adult speech, the frequency of tones in mono- and disyllabic words was noted in all seven mothers' speech. To provide adequate sampling, we used recordings made when the all infants were 13 months old. These will be used for comparison with the children's productions, as an estimate of tone distribution in the input.

Table 4.11. Tones in monosyllabic and disyllabic words of input speech

	Zuozuo	Keke	Xinyu	Yifan	Yiyi	Shi	Didi	Total
Mono	23	37	44	33	30	45	23	235
Tone1	5	10	10	11	7	12	5	60

Tone2	5	7	12	8	8	6	7	59
Tone3	7	13	8	6	6	13	6	58
Tone4	6	7	14	8	9	14	5	63
Di	73	89	70	83	81	94	51	541
T1-T0	2	10	7	4	4	10	3	40
T1-T1	4	5	2	6	6	7	7	37
T1-T2	3	2	2	3	2	4	4	20
T1-T3	3	3	2	3	2	1	2	16
T1-T4	0	3	3	1	1	3	2	13
T2-T0	8	8	12	3	8	10	3	52
T2-T1	1	1	2	1	1	2	1	9
T2-T2	1	0	1	0	3	0	0	5
T2-T3	5	4	3	5	5	5	2	29
T2-T4	4	8	5	4	2	3	6	32
T3-T0	3	7	4	8	6	9	5	42
T3-T1	4	3	2	3	1	3	3	19
T3-T2	2	6	0	5	2	2	0	17
T3-T4	5	2	4	5	4	6	1	27
T4-T0	8	8	7	10	12	12	6	63
T4-T1	5	3	1	2	6	5	2	24
T4-T2	8	6	7	8	4	2	3	38
T4-T3	3	2	3	5	6	6	0	25

T4-T4	4	8	3	7	6	4	1	33

The distribution of lexical tones in monosyllabic and disyllabic words in the speech of the seven mothers is shown in Table 4.11. Although there are individual differences between the seven mothers in their tones use in monosyllables, the overall distribution not greatly different among the mothers, with Tones 1, 2 and 4 accounting for about 20-33% each, and Tone 3 a little less frequent: on average, Tone 1 accounts for 25% (22-33%), Tone 2 for 23% (18-29%), Tone 3 for 25% (16-22%) and Tone 4 for 27% (19-32%) of all monosyllables.

Disyllables are more variable, and mothers produced more types than the children. In summary, the common tonal patterns used by mothers in disyllables were T4-T0 (12%), T2-T0 (10%), T3-T0 (8%), T1-T0 (7%), T1-T1 (7%) and T4-T2 (7%). We can see that the adult input speech is characterised by citation tones followed by the neutral tone.
Comparing the mothers' productions with the children's productions reveals that the proportion of tone use in monosyllables was different at two developmental points. As noted earlier, the children produced considerably more Tone 1 and Tone 4 than the mothers at the 4-word point. As the children's lexicon grew, the frequency of tones remains different from the mothers' at the 25-word point but is gradually coming to approximate it: Tone 1 is 33% for children and 25% for mothers; Tone 4 is 40% for children and 27% for mothers, and Tone2: 19% for children, 23% for mothers.

The situation is different when disyllables are considered. T1-T4 (25%) and T1-T0 (20%) account for all the frequently-used patterns at the 4wp which are not seen in adult language. The infants were approaching more adult-like tonal patterns by the 25-word point: T1-T0: 18% T4-T0: 12% T2-T0: 6%. For instance, for the two emerging patterns, the proportion of T2-T0 is 6% for the children, 10% for the mothers; Tone T2-T4 is 6% for both children mothers.

4.5 Discussion and conclusion

4.5.1 Primary findings of tone production in Mandarin-learning children

Mandarin infants' tone learning depends on prior experience in babbling and is characterised by developmental change between the early and late sessions that we analysed: 4-word point

and 25-word point. Tones were found to develop along a different timeline from segments. The analysis of segmental development (Ch. 3) generally supports Ferguson and Farwell's finding of relative accuracy in the first words. However, Mandarin children's tone use in their first words (4-word point) is not always accurate and is largely restricted to Tone 1 and Tone 4, although the other tones were also targeted. These patterns are in line with the strong preference we observed in babbling. Tone 2 is much less often produced but is the most accurate, and production of Tone 3 cannot be detected at this point. This limitation is likely to be due to both the perceptual salience of Tones 1 and 4 and the existence of well-controlled vocalizations supported by babbling practice. After practice with their production in babble, Tone 1 and Tone 4 are over-selected at the 4-word point; these tones become template-like patterns and are over-used as substitutes for the other tones. On the other hand, Tone 2 words begin to be selected and used accurately through individual episodes or perceptual experience with the pattern as a whole (item learning). Achieving the motor control for the articulation and planning of two or three pitch changes within a single syllable (as required for Tone 2 or Tone 3) could be challenging for the inexperienced infant and memory for the sequence may be challenging as well. Mandarin-learning children may pick up Tone 2 through frequent exposure to specific instances or exemplars rather than abstract sequences.

Similarly, in disyllables, the Tone 1 and Tone 4 combinations present in the children's production – T1-T0 and T1-T4, both of which involve the use of level tone followed by a drop – tend to be the most favored patterns, both in terms of total occurrence and the number of children who use them. T4-T4, T1-T1 and T4-T0 are also produced with some frequency. Among these frequent patterns, T4-T0, T1-T0 and T1-T1, which occur mainly in kinship term and onomatopoeia and which therefore are easily accessible through item learning, are accurate enough. However, all the T1-T4 and T4-T4 patterns involve substitution for other tones; there are no T1-T4 target words at all. The children were not able to successfully reproduce neutral tone; more than half of the uses of T1-T4 and T4-T4 are substitutes for the citation tone followed by neutral tone (e.g., T2-T0, T3-T0, T4-T0). The biological constraints reflected in the natural early preference for level and falling pitch (Tone 1 and Tone 4) are still shaping children's tone production at the 4-words point.

At the later lexical point (25-word point) the children are still producing Tone 1 and Tone 4 the most. Tone 2 increases in use and Tone 3 first appears at this point. Tone 2 and Tone 3 gradually come to be within infants' motoric capacity after the onset of regular word

production, presumably in response to longer exposure to the input language. After long practice in both babble and the early words, there is also progress in accuracy for Tone 1 and Tone 4. By the 25-word point these tones rarely serve as substitutes for other more difficult tones. After a period of 'accurate' item learning at the 4-word point, Tone 2 is extended to words with other target tones and is now mainly used to substitute for Tone 3. The limited emergent use of Tone 3 shows its item learning status with perfect matches to target.

In disyllables, T1-T0, T1-T4, T1-T1 and T4-T0 (listed in order of frequency, from highest to lowest) are still the most frequent patterns that the children produce, but they are not the only patterns. As the children learn more words, they produce more variable tonal patterns overall. With the increasing use of Tone 2 and the emergence of Tone 3 in monosyllables, Tone 2 and Tone 3 also emerge in disyllabic words at this point. The ambient language is now beginning to shape the children's phonological systems.

A high proportion of Tone 1 and Tone 4 is a characteristic of children's speech, as opposed to being input-related. According to CDI data and analyses of targets and actual adult input speech, the four tones are distributed in similar proportions in monosyllabic words. Adults' tonal patterns are more variable in disyllables. Children target fewer patterns than adults do. Although variable tones in words in the ambient language influence the early target tones, factors such as physiological constraints may bias young children's tone production. Individual differences were found in the extent to which the children succeeded in matching adult tonal patterns. These Mandarin children could not match tones appropriately at the 4-word point, but they generally conformed relatively closely to the adult pattern by the 25-word point.

4.5.2 Independent developmental trajectories of the four tones and neutral tone

Our findings are more or less in agreement with Zhu (2002), which found that the

> High level (Tone 1) and falling tones (Tone 4) were present at the beginning of the children's lexical use between the age of 1;2 and 1;4. Falling–rising tones (Tone 3) emerged last, between the age of 1;4 and 1;7. (p.188).

However, in our study the four lexical tones are each associated with a separate developmental path. Although the data here are limited, it seems plausible that perception-based and production-based factors interact to determine the tone patterns of early words in Mandarin-learning children.

Early tone use in words: Tone 1 and Tone 4

As we indicated in the results section, a preference for Tone 1 and Tone 4 is apparent in Mandarin-learning children's monosyllabic words across the two lexical points. It is possible that the physiological complexity of articulation influences tone production in early childhood. Tone is a function of the rate of vocal fold vibration (see Ohala, 1978, for a review). Changes in rate of vocal fold vibration are made by manipulating tension in the vocal folds, which is increased or decreased by the laryngeal muscles (Sagart, Halle, Boysson-Bardies, Arabia-Guidet 1986). Tone 4, with its falling pitch movement, is a natural speech production gesture and is primarily defined by relaxation of the laryngeal muscles. Tone 1 also requires less physiological effort; it only demands keeping the status of the laryngeal muscles stable/constant, with no further movement needed. This simplicity of execution can explain why Tone 1 and Tone 4 would be acquired earlier than Tone 2 (rising) and Tone 3 (falling–rising). Moreover, in an autosegmental account, Tone 1, as a level tone, consists of only one unit or feature, while contour tones can be considered a sequence of two or several different level tones: Tone 2 is a sequence of a low level and a high level tone; Tone 3 is a sequence of mid-level, low-level and high-level tones. Tse (1992) showed this in a study of Cantonese tone acquisition, which reported that among the nine tones in Cantonese (6 level tones and 3 contour tones), all the level tones were acquired earlier than any contour tones.

The early production of Tone 1 and Tone 4 might also be guided by acoustic salience. There are f0 differences between tones. The ranking by average f0 of the Mandarin lexical tones, ordered from high to low, is: Tone 1 ≈ Tone 4 >Tone 2> Tone 3 (Liu et al., 2007). Thus higher pitch might make Tone 1 and Tone 4 more salient in the input. It is found that phonological saliency can affect the acquisition of features. Components with higher phonological saliency would be acquired earlier than components with lower saliency (Zhu, 2002).

The first words are learned based on patterns already motorically and perceptually familiar from frequent production in the prelinguistic period (Vihman et al., 1985). In the course of babbling infants develop a range of vocal skills and maintain their early preferences for vocal forms when starting to say words. Auditory familiarity also paves the way for first word learning. Children pay more attention to adult patterns of the ambient language that match their own vocal forms (Vihman, 1996). In segmental studies particular consonant types, like

stop consonants, have been found to be highly frequent both in first words and in babble (Bernhardt & Stoel-Gammon, 1996). Similarly here, our study has found a strong similarity between babble and word production in the production of lexical tone. In chapter 2 we investigated the incidence of tone patterns in Mandarin-acquiring children's prelinguistic vocalization and found consistently high use of Tone 1 and Tone 4. The frequent occurrence of Tone 1 and Tone 4 in babbling gives children a good opportunity to practice them, and this serves as a basis for forming Tone 1 and Tone 4 words in first word production. So here children extend to meaningful words non-meaningful uses of Tone 1 and Tone 4, which are the most salient and most often practiced patterns in babbling.

The story of Tone 2

We see that there is a major shift in the use of Tone 2 between the two word-points in the infants who use it. At the 4-word point children's Tone 2 words, although few in number, typically reflect attempts at adult Tone 2 target words; in other words, they are highly accurate. Eighty percent of all three children's productions of Tone 2 are correct. With more prolific production at the 25-word point, an overall decrease in accuracy appears. Only 48% of the uses are correct; the remaining uses of Tone 2 are attempts at Tone 3.

Tone 2 demonstrates item learning and systematic development in the observational period of our study. Mandarin-acquiring children learn individual words with Tone 2 by straightforward imitation and associating it with particular contexts, then move to recognizing Tone 2 used in different situations. For example, Tone 2 is always used in the word *mau2* 'peek-a-boo' in the play routine and in the word *wai2* hello (on the phone) in making a pretend phone call. At the time when the child has already produced several first words, he/she begins to independently vary the Tone 2 pattern and learn to use Tone 2 as a substitute. Tone 2 at the earlier stage had been learned as an unanalyzed unit and involved a unique meaning (or a few such meanings). By the later stage, children have learned that Tone 2 can occur in other forms to give a range of different meanings. The acquisition of Tone 2 supports the idea that early representations are 'holistic' rather than made up of sublexical elements.

Moreover, at the 25-word point the children who used Tone 2 at the 4-word point are producing more Tone 2. It seems to be the case that once one or two words with Tone 2 have

been produced, practice with those leads children to focus on words with Tone 2 patterns in the input.

Why is Tone 3 the last to emerge?

Firstly, the late acquisition of Tone 3 may be due to articulatory difficulty. That is, Tone 3 demands greater motor sequencing and control of the rate of vocal cord vibration. Physiological complexity affects the precision of vowel and consonant productions in early childhood (see Smith, 2006); this may be the case for tone production as well.

Second, acoustic properties may play a role here, i.e. syllables that are acoustically more prominent are most likely to be reproduced. Tones have an intrinsic duration. Based on connected speech, the mean durations of the four tones are: T1 212ms, T2 201ms, T3 178ms, T4 195ms (Wang J., 1994). Apparently, T3 is the shortest tone in any medial position in connected speech, although that is not the case in isolation or sentence-final position. Meanwhile, the full-tone melody of Tone 3 (contour 214) usually occurs only when a monosyllabic target word is produced in isolation or at the end of an utterance. On other occasions, Tone 3 is not fully pronounced but instead is truncated, with the final fall omitted (contour 21: low falling). As a result, a Mandarin-learning child might hear the full Tone 3 produced considerably less frequently than any one of the other tones and could readily confuse it with Tone 4.

The extent of contour similarity between lexical tones varies. For example, Tones 2 and 3 exhibit similar F0 contours: both have a dynamic F0 shape (in comparison with Tones 1 and 3. which have dissimilar F0 contours: level vs. dynamic). The acoustic similarity between tones may be associated with the slower acquisition in Mandarin-learning children. Tones 2 and 3 have been reported to be more frequently misarticulated than Tones 1 and 4 by Mandarin-learning children aged 1; 6 to 3; 0 (Li & Thompson, 1977).

Tone 2 and Tone 3 are not only acoustically similar but are also related via phonological rules. Specifically, Tone 3 sandhi prescribes that the first syllable is substituted by Tone 2 when two successive Tone 3s are produced. In other words, in two syllables, Tone 3 is realised as Tone 2. In certain contexts Tone 2 can be considered an allotone of Tone 3 (Chen, 2000). Wang and Li (1967) and Peng (2000) have shown that listeners are not able to distinguish a Tone 3 that has undergone tone sandhi from a Tone 2. Therefore, applying a

tone sandhi rule reduces the occurrence of Tone 3 and increases the likelihood of confusion for young children.

Neutral tone in disyllables

Mandarin learners in our study apply equal stress to each syllable in the disyllabic word in the early stages rather than use neutral tone. This is in accord with the findings of Hua (2002), who reports that most of the errors children make involve treating weak syllables (neutral tone) as having citation tone (equal stress), especially in reduplication. That is, this is a kind of overstressing. In Hua's cross-sectional data only 57% of the youngest children (i.e., 12 out of 21, aged 1;6 - 2;0) attempt to produce weakly stressed syllables (neutral tone): see her Table 3.6. Of those attempts, only 5% of the children succeeded in accurately matching half the words with weak syllable (neutral tone). Also in Hua's longitudinal data all four children used citation tones for weak-stress syllables (neutral tone), especially for reduplicated forms. The weak stress (neutral tone) in reduplication took longer to stabilise than that of the other two types (e.g. Affix and Lexeme; also see her Table 4.10) and it had not stabilised by the end of the data collection (about age 2;0).

F0 in Mandarin functions to express both lexical tone and intonation; therefore, the stress pattern of words, either in isolation as word stress or in longer utterances as a nuclear pitch accent, is closely related to F0, with pitch as the most important correlate. Duration is also a primary cue for the recognition of neutral tone. The duration of neutral tone T0 is about 50%-60% that of the lexical tones (Lee et al., 2014). Children's late acquisition of neutral tone may relate to the physiological constraints, that is the limited motor control for prosody. Similar to English-learning children, who were found to reduce vowel production under weak stress (Post & Paine, 2018) or lengthened the final syllable (Vihman, Nakai & DePaolis, 2006) at the 4-word point, Mandarin-learning children find it hard to produce reduced syllables and therefore avoid producing neutral tones but apply citation tone (a kind of final-syllable lengthening).

It is also possible that caregivers exaggerate their speech in IDS (Li & Thompson, 1977) and accommodate to their children by imitating their patterns, so that neutral tone may not occur as frequently in children's input as in adult speech. Therefore, children may not receive

enough input containing neutral tone to allow them to learn it early on. However, caregivers' imitations were not included in our analysis and little is known about their frequency; this needs to be further explored.

5. DISCUSSION AND CONCLUSIONS

5.1 Introduction

This thesis has investigated the origin and development of the phonological system in Mandarin-learning infants, from babble to the early word period, in three respects: tones, segments and prosodic structures. Most existing studies of phonological development focus on European-language-acquiring children; as Mandarin is the most widely spoken tone language in the world, this study aims to fill the gap in understanding the phonological acquisition of Mandarin-learning infants. Lexical tone is an essential component of Mandarin phonology, but not of English. The investigation of the acquisition of lexical tone offers an insight into the development of a phonological system different to that which might be gained by studying the development of intonation, for example.

This concluding chapter summarises the major findings of the study, based on the longitudinal observational data, in relation to the research questions outlined in 1.4. The developmental patterns of Mandarin-learning children are discussed in relation to the theoretical framework.

5.2 Review of the research questions and major findings

Three major research questions have been addressed in this thesis:

1. To what extent do Mandarin lexical tone patterns occur in infants' babble vocalizations?
2. How is the phonology of the ambient language incorporated into Mandarin-learning children's early word production? What segmental advances and challenges will they have?
3. What tonal advances and challenges will they have in their early word production?

Four data-based analyses have been reported to answer these questions: a cross-linguistic comparison of tone production in 25 Mandarin- and English-learning infants aged 0;9 – 1;0 (babbling period); an analysis of prosodic and segmental development in the early word period (1;0 – 1;5) in 8 Mandarin-learning children; an analysis of the acquisition of tone in the early word period (1;0 – 1;5) of 8 Mandarin-learning children; and an analysis of

Mandarin adult input in relation to both segment and tone learning. Each chapter presents an independent research strand.

This thesis demonstrated several new findings regarding young Mandarin-learning children's phonological acquisition and word learning. It presents the interaction of different factors to shape children's patterns in development: biological constraints, language-specific features and individual child differences. And it investigates segment and tone acquisition as part of an emergent system.

In chapter 2, we aim to answer the question, when do Mandarin-learning children first produce the tone patterns of their native language? There are several lines of evidence that suggest that infants in the first year learn about their ambient language by listening to it (Werker & Tees, 1984; Kuhl, 2005; Jusczyk, 2002); therefore, they might show the influence of that ambient language in their own vocalizations. We have provided evidence based on tonal features that the ambient language already starts to have an impact in the first year of infants' vocal production. Significant differences in the use of tones were found in Mandarin- compared with English-learning infants' babble vocalizations. Some tonal patterns (e.g. Tone 1) occurred in the babbling of the Mandarin-learning infants but were absent from that of the infants from an English background. These data support Brown (1958)'s babbling drift hypothesis. The commonality seen in children learning different languages is due to their being constrained in similar ways; that is, they have the same physiological constraints on vocal production.

Chapters 3 and 4 have indicated some clear trends in segmental and tone development in the early word period of Mandarin-acquiring children. Although accuracy is an important measure of success in phonetic or word acquisition, here we have also presented the frequency of occurrence and change in each category from the child's perspective. This is because the child's challenge at this period is that of gaining not only the articulatory skills but also the competency to retain forms in memory and to access them any time; therefore, we aimed to assess Mandarin-learning infants' phonological acquisition in a systematic way.

Early word learning is established through previous experience of speech forms overall, including both others' input and the infant's own output. Mandarin-learning children's word

production starts with “context-limited” words (4-word point), which are elicited from specific exemplars in adult/child routines (e.g. *Mummy* [ma1ma0]; *Daddy* [pa4pa0]; *quack-quack* [ka1ka4]), rather than with the unfolding of a particular order of segmental contrasts (Jakobson, 1941/68). The word-forms at this word point are simple and underspecified because they lack phonetic neighbors. This is in accord with exemplar theory and the Linked-attractor models. Biological constraints characterize the word forms as well as tone at this moment for Mandarin-learning children. Accommodation to the ambient language is more salient within accessible phonetic categories (e.g. CV syllable; stop or nasal consonant; Tone 1 or Tone 4), that is, those already mastered within the babbling period.

Then there are progressive increases both in the number of categories and in complexity in terms of segments and tones. Through continued vocal practice, children progress to “context-flexible” sound-meaning pairs by generalizing accessible patterns or using their own patterns (or templates). Over time, infants’ word production become phonetically specified and more systematic. Emergentist phonology theory is applied to explain this development. The influence of the adult language also increases as children acquire a lexicon because knowing even a small set of words is useful in helping infants gain better representational access to the adult input. Having a limited range of output types simplifies access to representations as well as supporting on-line production while at the same time permitting more ambitious word choices.

5.3 Different approaches on Mandarin phonological acquisition

There is a need to discuss different approaches to studying Mandarin-learning children’s phonological acquisition. There are two relatively recent studies. A longitudinal study followed four Mandarin-learning children from 10 to 24 months in Beijing and showed early ages of tone acquisition (Zhu, 2002). Zhu found that tones were acquired first, earlier than consonants and vowels. The children produced the four Mandarin tones accurately in their spontaneous speech before age 2, and their production stabilized between ages 1;4 and 1;9. The age of tone stabilization was defined as the age at which children were able to produce the tones with an accuracy rate of 66.7% and maintain this accuracy rate or higher in all subsequent language samples (Zhu, 2002).

Zhu (2002) applied Jakobson's 'feature theory', markedness theory and a biological model to examine and explain the commonalities and differences between Mandarin- and other-language-learning children. Zhu (2002b) proposed a language-specific concept, 'phonological saliency', to explain the order of acquisition in Mandarin-learning children. That is, Mandarin-learning children are constrained in their phonological development by phonological saliency. Tones have the highest saliency because they are compulsory for every syllable and there are only four tones; change of tones results in a different lexical meaning.

This study found a similar order of emergence of the four tones as Zhu (2002) reported. For example, in Zhu (2002), Tone 1 and Tone 4 were present at the beginning of the children's lexical use, and Tone 3 emerged last. However, this study does not support Zhu (2002)'s proposal that the acquisition of tones was completed at early age. Appling usage-based theory or emergent phonology, we emphasize the developmental process per se, that is, the occurrence of tonal categories. Therefore, in our story each tone has its own developmental trajectory: Tone 1 and Tone 4 appeared first and seemed to be the most stable at the end of our investigation (at age 1;6). Tone 2 appeared later and less often than Tone 1 and Tone 4 and showed an "accurate" individual item-learning pattern at the 4-word point, but accuracy decreased with more prolific production at the end of this study (25-word point). This leaves open the question of how long this process will last. Tone 3 appeared last and was still relatively less often used at the end of this study. Taken together, it raises the issue when the four tones could be stabilised, which had been examined by Wong and her colleagues.

Wong and her colleagues (e.g. Wong, Schwartz & Jenkins, 2005; Wong, 2008; Wong, 2012) also investigated children's production of Mandarin tones in monosyllabic words, with a focus on the acoustic characteristics. These authors suggested that the acquisition of Mandarin lexical tone is a protracted process and is still not complete by age 3. The tones children produced were not the same as those of adults (adult-like), according to the combined results of both adult perception and acoustic analyses. In terms of the order of acquisition of the four tones, the acoustic analysis showed the order from most to least adult-like to be Tone 4, Tone 1, Tone 2, and Tone3. Perception tests showed that Tone 2 (70%) has similar perceptual accuracy to Tone 1 (78%) and Tone 4 (76%), while T3 received the lowest accuracy rates (44% correct out of all the 470 judgments made by the 10 judges). Children's tones that were identified as incorrect had different F0 contours and acoustic characteristics

from correct adult and child productions. Biological constraints and articulatory complexity were given as ways to interpret the findings. The ease of motor control for producing the four tones is a contributing factor to the order of tone acquisition.

Our data also support the late acquisition of some tones, as proposed in the studies of Wong and her colleagues. Wong and her colleagues studied the Mandarin tone acquisition of older children (e.g. 3 years old upwards), while this thesis investigated the development of tone in younger Mandarin-learning children. This thesis could be the prologue of Wong's studies, providing some ideas and evidence relating to the later acquisition of Tone 2 and Tone 3.

5.4 Phonological development from babbling to the early-word period

Here we consider the whole picture of Mandarin-learning children's early phonological development. First, infants practice the production of speech-like syllables for several months, from about 9 months on. Tone as an important feature of Mandarin phonology has proven to be difficult to distinguish and produce for non-tone-language-speaking children and adults (Wang et al., 1999), but Mandarin-learning children find no such difficulty in the early-word period. The study reported in Chapter 2 of this thesis was therefore designed to find out whether Mandarin-like tone can be identified in Mandarin-acquiring infants' babble vocalizations. Brown (1958)'s notion of "babbling drift" is supported by our study of Mandarin-acquiring infants' production, which illustrates the use of a large number of categorizable Mandarin-like tones, Tone 1 and Tone 4 in monosyllables, and combinations of these tones in disyllables. In contrast to Lee et al. (2017), who failed to identify an ambient language effect in the babbling of infants exposed to Mandarin, the ambient language effect could be identified in our cross-linguistically comparative study of tone patterns based on frequency of occurrence.

The babbling patterns of infants are also subject to simple biological constraints. The occurrence of falling tone (Mandarin Tone 4) is a feature of both English- and Mandarin-acquiring infants' babbling. The falling tone (or Tone 4) is articulated by decreasing vocal fold vibration or releasing the tension of the laryngeal muscles, which results in a natural drop in pitch. No exposure to lexical tone is needed for infants exposed to English to produce falling tone. Instead, the frequent occurrence of Tone 4 in English infants' babbling likely results from its natural occurrence in vocal production. Similar biological constraints are also

reflected in the fewer occurrences of Tone 2 and the absence of Tone 3 in Mandarin infants' babbling, for both tones require some degree of motor control.

Children continue to accumulate knowledge of certain phonetic gestures to prepare early words. McCune and Vihman (2001) have demonstrated that the attainment of VMS shows infants' growing ability to consistently access speech-like motoric patterns with limited cognitive resources. Our data shows Mandarin-exposed infants' babbling is characterised by emergent motor control over specific tones. Not all adult lexical tones are represented or represented equally. Frequent usage of Tone 1 and Tone 4, less frequent usage of Tone 2 and no use of Tone 3 was a common pattern among Mandarin-acquiring infants over the 4-month period. Therefore, infants' self-produced actions, along with the ambient input, drives infants' production towards the adult language pattern.

The next definable stage is word-learning, from about 12 months of age. In response to prolonged ambient-language exposure, as well as vocal practice, Mandarin children begin to produce words with recognizable segment and tone patterns. Biological constraints and ambient language shaping in the pre-linguistic period carry over to the word-learning period in the second year. Easily accessible word forms and tone patterns are observed in Mandarin children's first words. Like most early words, the word shapes of Mandarin-learning children's target words and the forms they produce are typically CV or CVCV and include mainly stops and nasals. This is in accord with the frame and content model (Davis & MacNeilage, 1995, 2000),

We have also provided evidence of phonological selectivity (MacNeilage, 1979; Vihman et al, 1985) and item learning (see Waterson 1971), which is a plausible way for a child to get into first-word use. Tone 1 or Tone 4 monosyllabic words and Tone 1 and Tone 4 combinations in disyllabic words are the most frequent patterns seen im Mandarin-learning children's first words. This supports the articulatory filter model (Vihman, 1993), which posits that children's attention to adult speech is influenced by patterns matched to their own articulatory output and establishes an auditory feedback loop. Tone 1 and Tone 4 are well-practised in babbling; their familiarity leads these patterns to be established within Mandarin-learning infants' first words or meaningful lexical items. The articulatory filter mediates the input in rendering words with such patterns salient. However, it is not possible to avoid less easily accessible tone patterns, since there are only four tone categories in Mandarin. Tone 2,

a tone less frequently practiced in babbling, is less commonly selected by Mandarin children and occurs in many often-used words in the action context of a familiar routine in this period (4wp in Chapter 4). For instance, *wei2* 'hello (on the phone)', *mau2* 'peek-a-boo', *kei2(ni0)* 'give/ here you are', *na2* 'take' are all examples of the communicative function. Item learning is also shown in words with fricatives or affricates, which require the kind of precise articulatory control that is less common in early words. Two such words produced by Mandarin-learning children occur in a restricted context: *tɕiɛ3tɕiɛ0* 'older sister' and *ɕiɛ4ɕiɛ0* 'thanks'.

The 'implicit selective' first words are therefore close enough to adult words. There are different stories for syllable length, segment and tone properties in Mandarin-learning children. Each of the children's forms closely matched the length of the adult forms in at least one token (mean: 93%). Tone 1 and Tone 4 words are highly similar to the adult target at the segmental level but less similar at the tone level (Tone 1: 43%; Tone 4: 50%). Tone 2 words are very accurate (80% match) as a whole, although they are less frequently used (5 or 15% occurrence). As it is known that pitch develops early, here we can see that easy Mandarin tones which require less motoric control (Tone 1 and Tone 4) are already explored at will by children at the first-word point, and therefore may be 'less accurate'. Note that all matches or determinations of accuracy in the thesis have been decided from the perspective of child performance, which means that Tone 1 and Tone 4 are over-produced by Mandarin children and therefore sometimes function as substitutes for Tone 2 or Tone 3, or for one another. If we examine this from the adult perspective, all uses of target Tone 4 and almost all uses of target Tone 1 (5 out of 6) are realized accurately.

In the next stage, after the accumulation of a small lexicon through early individual lexical learning, in cooperation with sustained phonetic practice and advances in articulation and phonological memory, Mandarin children start to access more words with more challenging segmental and tonal patterns. They extend their existing routine word patterns to new word-form targets.

In harmony with the findings of Ferguson and Farwell (1975), at this stage we found variability in the occurrence of forms and a regression in accuracy. More prosodic structures are used. Diphthongs in monosyllabic words (CVV), diphthongs in disyllabic words (CVVCVV), closed monosyllables (CVC) and consonantal variegation (C_1VC_2V) emerge at

this point.

Ambient language influence increases when children become more advanced in language use. Mandarin input (39%) is found to provide more exposure to fricatives and affricates than do other, more often studied languages. Over a period of lexical learning, these Mandarin children, who produced 8% fricatives or affricates at the 4wp, and 23% at the 25wp, increased their production towards the fricatives or affricates in response to frequency of occurrence in the input.

Increasing use of Tone 2 (21%) and decreasing use of Tone 1 (36%) and Tone 4 (40%) were found. Tone 3 emerged first in some children's production but showed the characteristics of 'item learning' - that is, highly accurate production of selected items. Mandarin-learning children produced more tonal combinations in disyllables at this later point and demonstrated increased complexity beyond the Tone 1 and Tone 4 combinations of the first-word period.

Meanwhile, the resultant regression in matching to targets was observed in both segments and tone. Mandarin-learning children have by this stage become more familiar with producing fricatives or affricates and Tone 2 but are now less accurate in using them. There is high use of diphthongs and triphthongs but this usage is inaccurate, as the acquisition of vowels is a slow process (Lieberman, 1980).

Based on the evidence of sufficient exposure to variegation in the input and more frequent use of Tone 2 and of fricatives or affricates, which require precise motor control, the central difficulty at this point for Mandarin-learning children, who attempted and produced the least variegated forms among the language groups available for comparison, is phonological memory rather than articulation. Phonological memory is the ability to remember or represent arbitrary word forms at will. Young children are not able to keep arbitrary form-meaning associations in memory and retrieve them for use at any time (Bloom, 2000; Vihman, 2014, chapter 2). Mandarin-learning children stand out by producing the most reduplicated patterns out of four language groups (Vihman, Ota, Keren-Portnoy, Lou & Choo, under review). However, there may be less motivation for Mandarin-learning children to achieve complex and precise forms due to the relatively simple language structure and the flexibility of IDS. The reduplicated forms may be sufficient for a caregiver to recognize children's intentions. In some cases, caregivers even accommodate to children's speech habits. In Mandarin IDS,

words have many acceptable forms, for example, the variegated word *$t^hu4tsi0$* 'bunny' could also be reduplicated *t^hu4 t^hu0*, or monosyllabic *t^hu4.* Many words can be produced with or without reduplication

The usage-based emergentist idea that 'language is learned through use' (Pierrehumbert, 2003a, 2003b; Bybee, 2001) has been applied to studies of phonological development, for example, in the linked-attractor model (Menn et al, 2013) and whole-word phonology (Vihman &Croft, 2007). Experience of producing words, even if inaccurately, enables the child to gain representational access to the adult word pattern, which supports the development of phonological memory for word patterns (Keren-Portnoy et al., 2010; Parra, Hoff & Core, 2011). Children's favoured patterns play an important role as templates for accommodation to the adult language. Production of truncated, harmonized or restructured words has been found in children learning many languages (Vihman, 2019). We supported this idea by finding Mandarin-leaning children's specific patterns. Three children either over-selected CV with a specific consonant or managed the challenge by constraining the first consonant of disyllabic words to their preferred consonant. One child demonstrated a CVC and another a $CV_{GLIDE}V$ pattern. Perhaps due to the simple structures of Mandarin, the templates we found in Mandarin-learning children are less like those analysed by Vihman (2019), whose study was based on whole prosodic structures, but more like those reported by Kehoe (2015), who tended to identify a word-initial consonant as marking a template. However, the possible templates are found only in segments, not in tones, in Mandarin-learning children. The Mandarin-learning children all used a similar strategy for tones, possibly due to the limited choice of tone categories. The prosodic pattern in children's templates also indicates which prosodic pattern of the input language can be easily accessed.

Children's first words are more 'phonetic'; the 'phonological' stage begins at their 25-word point. Despite adapting novel words to templatic patterns segmentally, children acquiring Mandarin start to pick up on the similarities between tone patterns and build a phonological system through a self-organizing process. Tone 3 is substituted by acoustically-related Tone 2 or Tone 4 rather than Tone 1. The children started to master disyllabic words with a 'trochaic tone pattern' (a citation tone followed by a neutral tone, e.g. Tone1-Tone0), producing a relatively shorter and lighter second tone for disyllables, which corresponds to the adult pattern. These reflect the first step in the construction of a phonological grammar. The overall picture is that children evolve from an initially impoverished system to a category-rich

system, both segmentally and tonally.

5.5 Implications and directions for future research

Information on the development trajectory observed in Mandarin-learning children will be particularly valuable for clinical assessment and treatment of phonological disorders in China. Information on the general patterns of phonological development in children growing up in different linguistic backgrounds can be used as a framework for speech- language pathologists, educators, and linguists working with children learning the same language in different parts of the world, such as children learning Mandarin in different parts of China and immigrant children learning Mandarin in the United Kingdom and elsewhere.

This study enriches our understanding of the early period of phonological acquisition in Mandarin-learning children. Our data indicate the late acquisition of some tones (Tone 2 and Tone 3), which in line with the results of Wong and her colleagues, who studied Mandarin tone acquisition of children 3 years old and upwards. It leaves unclear how Mandarin infants use these tones from age 18 months to 3 years old. It will be important for future work to examine the process.

REFERENCES

Bernhardt, B. H. and Stemberger, J.P. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Bernhardt, B. & Stoel-Gammon, C. (1996). *Underspecification and markedness in normal and disordered phonological development*. In Johnson, C. E. & Gilbert, J. H. V. (eds), *Children's language*, vol. 9, 33–54. Mahwah NJ: Lawrence Erlbaum Associates.

Behrens, H. (2006) The input–output relationship in first language acquisition, Language and Cognitive Processes, 21:1-3, 2-24.

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Boysson-Bardies, B. (1993). *Ontogeny of language-specific syllabic productions*. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *NATO ASI series D: Behavioural and social sciences, Vol. 69. Developmental neurocognition: Speech and face processing in the first year of life* (p. 353–363). Kluwer Academic/Plenum Publishers.

Boysson-Bardies, B., Sagart, L. and Durand, C., (1984). Discernible differences in the babbling of infants according to target language. *Journal of child language*, *11*(01), pp.1-15.

Boysson-Bardies, B., Hallé, P., Sagart, L. and Durand, C., (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of child language*, *16*(01), pp.1-17.

Boysson-Bardies, B. and Vihman, M.M., (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67, pp.297-319.

Bolinger, D. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford: Stanford University Press.

Bohm, L.A., Nelson, M. E., Driver, L. E., Green, G.E. (2010). Babbling, Vegetative Function, and Language Development after Cricotracheal Resection in Aphonic Children. *The Laryngoscope*, 120(12), 2494 – 2497.

Brown, R. (1958). *Words and Things*. Glencoe, Illinois: Free Press.

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* (pp. 259–280). Amsterdam: John Benjamins (Series: Language Learning and Language Teaching).

Burton-Roberts, N., Carr, P. & Docherty, G. (2000). Introduction. In N. Burton-Roberts, P. Carr & G. Docherty (eds.), *Phonological knowledge: Conceptual and empirical issues* (pp.39-66). Oxford: Oxford University Press.

Bybee, J. & Hopper, P. (2001). Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Casillas, M., Brown, P., Levinson S.C. (2019). Early Language Experience in a Tseltal Mayan Village. *Child Dev*. Dec 31.

Cao, J. F. (1986). An analysis of Mandarin neutral tone syllables (in Chinese). Applied Acoustics (in Chinese), 4, 3-8.

Chao, Y. R. (1930/1980). A system of tone letters. Le Maître Phonétique 45:24–27. *Repr., Fangyan* 1980(2),81-83.

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

Chen, L.M. and Kent, R.D. (2009). Development of prosodic patterns in Mandarin-learning infants. *Journal of Child Language*, *36*(01), 73-84.

Chen, M. (2000). *Tone sandhi: Patterns across the Chinese dialects*. Cambridge: Cambridge University Press.

Chen, T.-M., Dell, G. S., & Chen, J.-Y. (2004). *A cross-linguistic study of phonological units: Syllables emerge from the statistics of Mandarin Chinese, but not from the statistics of English*. Presented at the 26th Annual Meeting of the Cognitive Science Society, August 5-7, Chicago.

Chen, Y. & Xu, Y. (2006). Production of weak elements in speech—evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, 63, 47-75.

Cheng, C.C. (1973). *A synchronic phonology of Mandarin Chinese* (Vol. 4). Walter de Gruyter.

Cheng, C.M. (1982). ANALYSIS OF PRESENT-DAY MANDARIN/今日汉语的统计分析. *Journal of Chinese Linguistics*, pp.281-358.

Chomsky, N. (1964). *Discussion. Monographs of the Society for Research in Child Development*, 29, no. 1, The Acquisition of Language: Report of the Fourth Conference Sponsored by the Committee on Intellective Processes (pp. 35-42).

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Clopper, C. G. & Turnbull, R. (2018). *Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors*. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler & M. Zellers (eds.), *Rethinking reduction*, 25–72. Berlin: de Gruyter Mouton.

Clumeck, H. (1977). Topics in the Acquisition of Mandarin Phonology: A Case Study. Papers and Reports on Child Language Development, 13, 37-72.

Clumeck H. (1980). "*The acquisition of tone*," in Child Phonology: Production Vol. I eds Yeni-Komshian G. H., Kavanagh J. E., Ferguson C. A. (New York, NY: Academic Press), 257–275.

Cruttenden, A. (1994). *"Phonetic and prosodic aspects of Baby Talk," in Input and Interaction in Language Acquisition,* eds C. Gallaway, and B. J. Richards (New York, NY: Cambridge University Press), 135–152.

Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Cutler, A. & Chen, H-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165-179.

Da, J. (2007). Chinese text computing corpus. http://lingua.mtsu.edu/chinese-computing/copyright.html.

Davis, B.L. & MacNeilage, P.F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech and Hearing Research*, 33, 16–27.

Davis, B.L. & MacNeilage, P.F. (1995). The articulatory basis of babbling.*Journal of Speech, Language, and Hearing Research*, *38*(6), pp.1199-1211.

Davis, B. L. & MacNeilage, P. F. (2000). An embodiment perspective on the acquisition of speech perception. *Phonetica*, 57, 229-241.

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, 75-107.

Delattre, P. (1961). La leçon d'intonation de Simone de Beauvoir, étude d'intonation déclarative compare. *The French Review*, 35, pp. 59–67.

Demuth, K., & Fee, J.E. (1995). *Minimal prosodic words in early phonological development*. Manuscript submitted for publication.

DePaolis, R., Vihman, M. M. & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development, 34,* 590-601.

Deschamps, I., Baum, S. R., Gracco, V.L. (2015). Phonological processing in speech perception: What do sonority differences tell us? *Brain Lang*, 149, 77-83.

DePaolis, R. A, Keren-Portnoy, T. & Vihman, M. M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology*, 7, 715.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, 19, 603–615.

Dreher, J. J., & Lee, P. C. (1966). *Instrumental investigation of single and paired Mandarin tonemes (Research Communication No. 13)*. Huntington Beach, CA: Advanced Research Laboratory, Douglas Aircraft Company

Dreher, J.J. and Lee, P.C.E. (1968). Instrumental investigation of single and paired Mandarin tonemes. *Monumenta serica*, pp.343-373.

Duanmu, S. (2000). *The phonology of standard Chinese*. New York: Oxford University Press.

Duanmu, S. (2002). *Phonology of Standard Chinese*. Oxford: Oxford University Press.

Eimas, P.D. and Miller, J. L. (1980). Contextual effects in infant speech perception. *Science*. 209, 1140–1141.

Edelman, G. (1989). *The Remembered Present: A biological theory of consciousness*. New York: Basic Books.

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, 421– 436.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.

Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Lang*. 25, 219–234.

Engstrand, O., Williams, K., & Lacerda, F. (2003). Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12-and 18-montholds. *Phonetica*, 60, 17-44.

Fernald, A. (1992). *Meaningful melodies in mothers' speech to infants*. In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Studies in emotion and social interaction. Nonverbal vocal communication: Comparative and developmental approaches* (p. 262–282). Cambridge University Press; Editions de la Maison des Sciences de l'Homme.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, 51, 419-439.

Ferguson C. A. (1986). *Discovering sound units and constructing sound systems: It's child play*. In J.S. Perkell and D.H. Klatt (eds.), *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Lawrence Erlbaum.

Fikkert, P. (1994). *On the Acquisition of Prosodic Structure*. Amsterdam: Holland Institute of Generative Linguistics.

Fontan L., Tardieu J., Gaillard P., Woisard V., & Ruiz R. (2015). Relationship between speech intelligibility and speech comprehension in babble noise. *Journal of Speech, Language, and Hearing Research*, 58, 977–986.

Gleitman, L. R. & Wanner, E. (1982). *Language acquisition: The state of the state of the art*. In E. Wanner & L. R. Gleitman (eds.), *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press.

Goad, H. and Ingram, D. (1987). Individual variation and its relevance to a theory of phonological acquisition. *Journal of Child Language*, *14*(03), pp.419-432.

Goldsmith, J. A. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), 8030–8035.

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, 15(4), 362–391.

Graf Estes K, Evans JL, (2007). Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci*, 18, 254–260.

Grieser, D.L. and Kuhl, P.K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental psychology*, *24*(1), p.14.

Gros-Louis, J., West, M. J, Goldstein, M. H. & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30, 509-516.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation. Sound files.* Cambridge: Cambridge University.

Hallé, P.A., De Boysson-Bardies, B. and Vihman, M.M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, *34*(4), pp.299-318.

Hao, M., Shu, H., Xing, A. & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods*, 40, 728–733.

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, 110(8), 581-616.

Hart, B. (1991). Input frequency and children's first words. *First Language*, 11, 289-300.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*, 20(2), 241–251.

Hou, J. Y. (2002). *The Outline of Modern Chinese Dialects*. Shanghai: Shanghai Education Publishing House.

Hsu, J.H. (1987). A study of the various stages of development and acquisition of Mandarin Chinese by children in Chinese milieu. *NSC report*.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals*. (A. R. Keiler, trans.). The Hague, Netherlands: Mouton.

Jeng, H.-H., (1979). The acquisition of Chinese phonology in relation to Jakobson's law of irreversible solidarity. *Proceedings of the 9th International Congress of Phonetic Sciences 2*, 155–61.

Johnson, K. (1997). *Speech perception without speaker normalization: an exemplar model*. In K. Johnson, & J. W. Mullennix (eds.), *Talker variability in speech processing* (pp. 145-166). San Diego: Academic Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44(4), 548–567.

Jusczyk, P. W. (1992). *Developing phonological categories from the speech signal*. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological Development: Models, research, implications* (pp. 17-64). Timonium, MD: York Press.

Jusczyk, P.W., Cutler, A., Redanz, N.J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Dev*, 64(3), 675-87.

Jusczyk, P.W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.
Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38, 338–350.
Kehoe, M. (2015). Lexical-phonological interactions in bilingual children. *First Language*, 35, 93-125.

Kent, R.D. and Murray, A.D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*, *72*(2), 353-365.

Kent, R.D. and Bauer, H.R. (1985). Vocalizations of one-year-olds. *Journal of Child Language*, *12*(03), pp.491-526.

Kent, R.D., Mitchell, P.R., Sancier M. (1991). *Evidence and role of rhythmic organization in early vocal development in human infants*. J. Fagard, P.H. Wolff (Eds.), *The development of timing control and temporal organization in coordinated action*. Elsevier Science Publishers, Amsterdam, pp. 135-149.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*. Timonium, MD: York Press.

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, 1280-1293.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606-608.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-843.

Kuhl, P.K. (2011). *Social mechanisms in early language acquisition: Understanding integrated brain systems supporting language*. In: Decety J, Cacioppo J, editors. *The handbook of social neuroscience* (pp. 649–667). Oxford UK: Oxford University Press.

Lahrouchi, M., & Kern, S. (2018). From babbling to first words in Tashlhiyt language acquisition: longitudinal two-case studies. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 63(4), 493-526.

Laufer, M.Z. (1980). *Temporal regularity in pre-speech*. In T. Murry & A. Murry (eds), *Infant Communication: Cry and Early Speech*. Houston: College Hill Press.

Legate, J. A. & Yang, C. (2007). Morphosyntactic learning and the development of tense. Language Acquisition, 14 (3), 315-344.

Lee, J-Y. (2003). Tense and length in English phonology. English Language and Literature 49, 1093–1144.

Lee, S.A., Davis, B. & Macneilage P. (2010). Universal production patterns and ambient language influences in babbling: a cross-linguistic study of Korean- and English-learning infants. *J Child Lang*, 37(2):293-318.

Lee, C-Y., Lekich, A. &Yu Zhang,Y. (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, 135, 1607.

Lee, C-C., Jhang, Y., Chen, L., Relyea, B. & Oller, D. K. (2017). Subtlety of ambient-language effects in babbling: A study of Englishand Chinese-learning infants at 8, 10, and 12 months. *Language Learning and Development*, 13, 100-126.

Li, P.J.K. (1977). Child language acquisition of Mandarin phonology. In *Proceedings of the symposium on Chinese linguistics* (pp. 295-316).
Li, C.N. & Thompson, S.A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185-199.

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

Li, A., Gao, J., Jia, Y., & Wang, Y. (2014). Pitch and duration as cues in perception of neutral tone under different Contexts in Standard Chinese. In Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA) (pp. 1-6). IEEE.

Lieberman P. (1980). *On the development of vowel production in young children*. In: Yeni-Komshian GH, Kavanagh JS, Ferguson CA, editors. *Child Phonology, Volume 1: Production*. New York: Academic Press.

Lin, S.-C. (1971). Phonetic development of Chinese infants. *Acta Psychologica Taiwanica Taipei, Chinese Psychological Association,* 13, 191–95.

Lin, T. (2001). Tantao Beijinghua qingyin xingzhi de chubu shiyan: Discussing preliminary experiments on the nature of non-stress in Pekingese (in Chinese). In Lin Tao *yuyanxue lunwen ji (in Chinese),* ed.by Lin Tao, 120–141. Beijing: Shangwu yinshuguan.

Lin, Y.H. (2007). *The Sounds of Chinese with Audio CD* (Vol. 1). Cambridge University Press.
Liu, H.M., Tsao, F.M., Kuhl, P.K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43, 912–917.

Liu L. & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, 133, 385–394.

Liu S, Samuel AG. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Lang Speech*, 47(2), 109-38.

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

Locke, J.L. and Pearson, D.M. (1992). Vocal learning and the emergence of phonological capacity: A neurobiological approach. *Phonological development: Models, research, implications*, pp.91-129.

Logan, G.D. (1988). Towards an instance theory of automatization. *Psychological Review*, 95, 492-527.

Lou, S., Vihman, M. M. & Keren-Portnoy, T. (2018). *Pitch differences in the babbling of Mandarinand English-learning infants*. Poster presented at the British Association of Academic Phoneticians. Kent.

Lu Y. (1995). *Putonghua de qingsheng he erhua: The weak stress and rhotacisation of Putonghua*. Beijing: Shangwu Press.

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*, 49, 11-49.

Macken, M. A. (1995). *Phonological acquisition*. In J. A. Goldsmith (ed.), *The Handbook of Phonological Theory*, pp. 671-696. Oxford: Blackwell.

MacNeilage, P. F. (1979). *Speech production*. Report to the Ninth International Congress of Phonetic Sciences, Copenhagen.

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. Hillsdale*, N.J.: Lawrence Erlbaum Associates.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465), 527–531.

Majorano, M., Vihman, M. M. & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, 10, 179-204.

Mattock, K. & Burnham, D. (2006). Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization. *Infancy*, 10(3), 241-265.

Mattock, K., Molnar, M., Polka, L. & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, 106, 1367–1381.

Maye, J., Weiss, D.J., Aslin R.N. (2008). Statistical phonetic learning in infants: facilitation and feature generlization. *Dev Sci*, 11, 122–134.

Mintz, T. H. (2006). *Finding the verbs: Distributional cues to categories available to young learners*. In: Hirsh-Pasek K, Golinkoff RM, eds. *Action Meets Word: How Children Learn Verbs*. (31–63) New York: Oxford University Press.

McCarthy, J. (1979). *Formal problems in Semitic phonology and morphology*. PhD dissertation. Cambridge, MA, MIT.

McCune, L. & Vihman, M. M. (1987). Vocal Motor Schemes. *Papers and Reports on Child Language Development*, *26*, 72-79.

McCune, L. and Vihman, M.M. (2001). Early Phonetic and Lexical Development: A Productivity Approach. *Journal of Speech, Language, and Hearing Research*, *44*(3), 670-684.

McGillion, M., Herbert, J. S., Pine, J., Vihman, M. M., DePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development*, 88(1), 156–166.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. and Amiel-Tison, C., 1988. A precursor of language acquisition in young infants. *Cognition*, *29*(2), pp.143-178.

Mehler, J., Dupoux, E., Nazzi, T. & Dehaene-Lambertz, G. (1996). *Coping with linguistic diversity: The infant's viewpoint*. In J. Morgan & K. Demuth (eds.), *Signal to Syntax*, pp. 101-116. Mahwah, NJ: Lawrence Erlbaum.

Menn, L. (1971). Phonotactic rules in beginning speech: A study in the development of English discourse. *Lingua*, 26, 225-251.

Menn, L. (1983). *Development of articulatory, phonetic, and phonological capabilities*. In B. Butterworth (ed.), *Language Production, vol. 2. London: Academic Press*.

Menn, L., Schmidt, E. & Nicholas, B. (2009). Conspiracy and sabotage in the acquisition of phonology: dense data undermine existing theories, provide scaffolding for a new one. *Language Sciences*, 31(s 2–3), 285–304.

Menn, L. & Vihman, M. M. (2011). Features in child phonology: inherent, emergent, or artefacts of analysis? In N. Clements & R. Ridouane (eds.), *Where Do Phonological Features Come From? The nature and sources of phonological primitives*. (Language Faculty and Beyond 6.) (pp. 261-301).

Menn, L., Schmidt, E., & Nicholas, B. (2013). *Challenges to theories, charges to a model: the Linked Attractor model of phonological development*. In M. Vihman & T. Keren-Portnoy (eds.) *The emergence of phonology: whole-word approaches and cross-linguistic evidence, CUP*.

Messum, P. & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, 53, 125-140.

Michèle Pettinato, M., De Clack, I., Verhoeven, J., Gillis, S. (2015). The production of word stress in babbles and early words: a comparison between normally hearing infants and infants with cochlear implants. Glasgow: ICPHS Conference Paper.

Moeller, M.P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., Wood, S., Lewis, D., Pittman, A., Stelmachowicz, P. (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I--phonetic development. *Ear Hear*, 28(5), 605-27.

Moon, C., Cooper, R.P. and Fifer, W.P., (1993). Two-day-olds prefer their native language. *Infant behavior and development*, *16*(4), 495-500.

Morgan, J. L. & Demuth, K. (eds.) (1996). *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, L. & Wren, Y. (2018). A Systematic Review of the Literature on Early Vocalizations and Babbling Patterns in Young Children. *Communication Disorders Quarterly*, 40(6), 152574011876021.

Munson, B. (2001). Phonological Pattern Frequency and Speech Production in Adults and Children. *Journal of Speech Language and Hearing Research*, 44(4), 778-92.

Newhoff, L.B.L.M. and Mesalam, L. (1980). Individual differences in early child phonology. *Applied Psycholinguistics*, *1*(1), pp.7-30.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
Ohala, J. J. (1978). *Production of tone*. In Tone: A Linguistic Survey, V.A. Fromkin (ed.) (Academic Press, NY), pp. 5-32.

Oller, D. K. (1980). *The emergence of the sounds of speech in infancy*. In G. YeniKomshian, J. F. Kavanagh & C. A. Ferguson (eds.), *Child Phonology, I: Production*. New York: Academic Press.

Oller, D.K. and Eilers, R.E., (1982). Similarity of babbling in Spanish-and English-learning babies. *Journal of Child Language*, *9*(03), 565-577.

Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. *Journal of Speech & Hearing Research*, 28(1), 47–63.
Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59(2), 441–449.

Oller, D.K., Wieman, L.A., Doyle, W.J. and Ross, C. (1976). Infant babbling and speech. *Journal of child language*, *3*(01), pp.1-11.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum.

Parra, M., Hoff, E. & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish–English bilingually-developing two-year-olds. *Journal of Experimental Child Psychology*, 108 (1), 113–25.

Pater, J. (1997). Minimal Violation and Phonological Development. *Language Acquisition*, 6(3), 201-253.

Peng, S. (2000). *Lexical versus 'phonological' representations of Mandarin Sandhi Tones*. In Broe M. and Pierrehumbert J. (eds.), *Papers in laboratory phonology 5: acquisition and the lexicon*. pp. 152-167. Cambridge: Cambridge University Press.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club.

Pierrehumbert, J. (2001). The phonetic grounding of phonology, *Bulletin de la Communication Parlee*, 5, 7-23.

Pierrehumbert, J. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115-154.

Pierrehumbert, J. (2003b). *Probabilistic theories of phonology*. In *Probability Theory in Linguistics*, RensBod, Jennifer B. Hay, and Stefanie Jannedy (eds.), 177– 228. Cambridge, MA: MIT Press.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2, 33–52.
Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221–231.

Polka, L. & Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. *J. Exp. Psychol. Hum. Percept. Perform.,* 20(2), 421-435.
Priestly, T.M.S. (1977). One idiosyncratic strategy in the acquisition of phonology. Journal of Child Language, 4, 45-65.

Quam, C., & Swingley, D. (2010). Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, 62, 135-150.

Renner, L. (2017). *The Magic of Matching: Speech production and perception in language acquisition*. PhD thesis, Stockholm University.

Robb, M. P., Saxman, J. H., & Grant, A. A. (1989). Vocal fundamental frequency characteristics during the first two years of life. *Journal of the Acoustical Society of America, 85*(4), 1708–1717.

Robb, M. P., & Tyler, A. A. (1995). Durations of young children's word and nonword vocalizations. *Journal of the Acoustical Society of America*, 98, 1348–1354.

Rose, Y. (2018). *Phonological categories and their manifestation in child phonology*. In M. Hickman, E. Veneziano & H. Jisa (eds.), *Sources of Variation in First Language Acquisition: Languages, contexts, and learners. Trends in language acquisition research (TiLAR)*. Amsterdam: John Benjamins.

Saffran, J.R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, 274, 1926-1928.

Sagart, L., Halle, P., de Boysson-Bardies, B., and Arabia-Guidet, C. (1986). Tone production in Modern Standard Chinese: An electromygraphic investigation. *Cahiers Linguistique Asie-Orientale*, 15, 205-221.

Scheer, T. (2012). *Direct Interface and One-Channel Translation. Vol.2 of A Lateral Theory of phonology*. Berlin: Mouton de Gruyter.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11, 563-605.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure. Cambridge*. MA: The MIT Press.

Shen, X. S. (1989). *Prosody of Mandarin Chinese*. Berkeley: University of California Press.

Shi, R. (2010). Contextual variability and infants' perception of tonal categories. *Chin. J. Phon*. 2, 1–9.

Shiu, H.S. (1990). *The phonological acquisition by Mandarin-speaking children: A longitudinal case study on children from nine months through three years old*. National Taiwan Normal University.

Singh, R. (2005). *Reduplication in Modern Hindi and the theory of reduplication*. In B. Hurch (ed.), *Studies on Reduplication*, pp. 263-282. Berlin: Mouton de Gruyter.

Singh, L. & Fu, C.S. L. (2016). A New View of Language Development: The Acquisition of Lexical Tone. *Child Dev*, 87(3), 834-54.

Smith, N.V. (1973). *The Acquisition of Phonology: A Case Study*. Cambridge: The University Press.

Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39, 331–349.

Snow, C. E. (1972). Mothers' speech to children learning language. *Child Dev*. 1972, 549–565.

Stamp, D. (1979). A Dissertation on Natural Phonology. Bloomington: IULC.

Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: A longitudinal study. *Journal of Speech & Hearing Research*, 28(4), 505–512.
Stoel-Gammon, C. (1992). *Prelinguistic vocal development: Measurement and predictions*. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press.

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

Stoel-Gammon, C. (1998). Role of babbling and phonology in early linguistic development. *Transitions in prelinguistic communication*, *7*, 87-110.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language*, 38, 1-34.

Su, A.T. (1985). The acquisition of Mandarin phonology by Taiwanese Children. *Unpublished MA thesis, Fu Jen Catholic University*.

Sun, C. (2006). *Chinese: A linguistic introduction*. Cambridge University Press.

Thelen, E. (1991). *Motor aspects of emergent speech: A dynamic approach*. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 339–362). Hillsdale, NJ: Erlbaum.

Thelen, E., & Smith, L. B. (1994). *MIT Press/Bradford book series in cognitive psychology.A dynamic systems approach to the development of cognition and action*. The MIT Press.

Thelen, E., Corbetta, D. and Spencer, J.P., (1996). Development of reaching during the first year: role of movement speed. *Journal of experimental psychology: human perception and performance*, *22*(5), 1059.

Thevenin, D. M, Eilers, R. E, Oller, D. K, & Lavoie, L. (1985). Where's the drift in babbling drift? A cross-linguistic study. *Applied Psycholinguistics*, 6, 3-15.
Tsao, F. M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chin. J. Psychol*, 50, 111–124.

Tse, A. C.-Y. (1992) *The acquisition process of Cantonese phonology: A case study*. M.Phil. thesis, University of Hong Kong.

Vihman, M. M. (1976). From pre-speech to speech: on early phonology. *Stanford Papers and Reports on Child Language Development*, 12, 230-244.

Vihman, M.M. (1978). Consonant harmony: Its scope and function in child language. *Phonology*, 2, 281-334.

Vihman, M.M., Macken, M.A., Miller, R., Simmons, H. & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, 397-445.

Vihman, M.M., Ferguson, C.A. & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*, *7*(01), pp.3-40.

Vihman, M. M. & Greenlee, M. (1987). Individual differences in phonological development: Ages one and three years. *Journal of Speech and Hearing Research*, 30,503-521.

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*. New York: Academic Press.

Vihman, M. M. & Velleman, S. (1989). Phonological reorganization: a case study. *Language and Speech*, 32, 149-170.

Vihman, M. M. & Velleman, S. L. (2000). *Phonetics and the origins of phonology*. In N. Burton-Roberts, P. Carr & G. Docherty (eds.), *Phonological knowledge: Its nature and status* (pp. 305-339). Oxford: Oxford University Press.

Vihman, M.M. (1992). Early syllables and the construction of phonology.*Phonological development: Models, research, implications*, pp.393-422.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21(1-2), 61–82.

Vihman, M. M. & McCune, L. (1994). When is a word a word? *Journal of Child Language*, 21, 517-542.

Vihman, M.M., Kay, E., de Boysson-Bardies, B., Durand, C. and Sundberg, U. (1994). External sources of individual differences? A cross-linguistic analysis of the phonetics of mothers' speech to 1-yr-old children. *Developmental Psychology*, *30*(5), 651.

Vihman, M. M. (1996). *Phonological Development: The origins of language in the child*. Oxford: Basil Blackwell.

Vihman, M. M., & DePaolis, R. A. (1998). *Perception and production in early vocal development: Evidence from the acquisition of accent*. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Chicago Linguistic Society 34, Part 2: Papers from the panels* (pp. 373–386). Chicago, IL: CLS.

Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50, 336353.

Vihman, M.M., Nakai, S. & DePaolis, R.A. (2006). *Getting the rhythm righ : A cross-linguistic study of segmental duration in babbling and first words*. In: Goldstein, Louis,

Whalen, David and Best, Catherine T., (eds.) *Laboratory Phonology 8*. Mouton de Gruyter, New York, pp. 341-366.

Vihman, M. M. & Croft, W. (2007). Phonological development: Toward a 'radical' templatic phonology. *Linguistics*, 45, 683-725.

Vihman, M. M. (2009). *Word learning and the origins of phonological system*. S. Foster Cohen (ed.), *Advances in language acquisition*. Luton: Macmillan.

Vihman, M. M. (2014). *Phonological Development: The first two years. (2nd ed.)* Malden, MA: Wiley-Blackwell.

Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development. Bilingualism. *Language and Cognition*, 19, 69-88.

Vihman, M. M. (2017). Learning words and learning sounds: Advances in language development. *British Journal of Psychology*, 108, 1-27.

Vihman, M. M. & Majorano, M. (2017). The role of geminates in infants' early words and word-form recognition. *Journal of Child Language*, 44, 158-184.

Vihman, M. M. (2018). *First word learning*. In S-A. Rueschemeyer & G. Gaskell (eds.), *Oxford Handbook of Psycholinguistics*, 2 ed. Oxford: Oxford University Press.

Vihman, M., & Wauquier, S. (2018). *Templates in child language. In Sources of variation in first language acquisition: Languages, contexts, and learners* (Trends in Language Acquisition Research (TiLAR)). Amsterdam: John Benjamins.

Vihman, M.M. (2019). *Phonological Templates in Development.* Oxford: Oxford University Press.

Ward, G. & Hirschberg, J. (1985). Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation. *Language*, 61(4), 747-776.

Wang, B., Lu, S., Yang, Y. (2002). 汉语语句中重读音节音高变化模式研究 The pitch movement of stressed syllable in Chinese sentences (in Chinese). Acta Acustica (in Chinese), 27, 234–240.

Wang, J. (1994). *Syllable duration in Mandarin*. Proc. of the Fifth International Conference on Speech Science and Technology, pp.322-327,

Wang, N.M., Wu, C.M. and Kirk, K.I. (2010). Lexical effects on spoken word recognition performance among Mandarin-speaking children with normal hearing and cochlear implants. *International journal of pediatric otorhinolaryngology*, *74* (8), 883-890.

Wang, W. S-Y. & Li, K. P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, 10, 629-236.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106(6), 3649-3658.

Waterson, N. (1971). Child Phonology: A Prosodic View. *Journal of Linguistics*, 7, 179-211.

Werker, J.F. & Tees, R.C. (1984). Cross-language speech-perception – evidence for perceptual reorganization during the 1st year of life. *Infant Behav. Dev.*, 7(1), 49-63.

Werker, J.F., Cohen, L.B., Lloyd, V.L. (1998). Casasola M, Stager CL. Acquisition of word-object associations by 14- month-old infants. *Dev Psychol*, 34, 1289–1309.

Westermann, G. & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. Brain and Language, 89(2), 393-400.

Whalen, D.H., Levitt, A.G. and Wang, Q. (1991). Intonational differences between the reduplicative babbling of French-and English-learning infants. *Journal of Child Language*, *18*(03), 501-516.

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

Wong, P. (2012). Monosyllabic Mandarin tone productions by 3-year-olds growing up in Taiwan and in the United States: Interjudge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research,* 55(5), 1423-1437.
Wong, P. (2013). Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *J. Acoust. Soc. Am.*, 133, 434–443.

Xiandai Hanyu Cidian (1979). Modern Chinese Dictionary. Beijing: Shangwu yinshu guan.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

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.

Yip, M.J., (1980). *The tonal phonology of Chinese* (Doctoral dissertation, Massachusetts Institute of Technology).

Yip, M. (2002). *Tone*. (Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press.

Zhang, J. L., Lu, S. N. & Qi, S. Q. (1982). A cluster analysis of the perceptual features of Chinese speech sounds. *Chinese Linguistics*, 10, 190-206.

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

Zhu, Z. (2002). Phonological development in specific contexts: Studies of Chinese-speaking children (Vol. 3). Clevedon: Multilingual Matters Ltd.

APPENDIX

Appendix I. Occurrence of tones in North China Mandarin-acquiring children's disyllabic vocalizations (1st= first syllable, 2nd= second syllable; bold face indicates the patterns meet the criteria). The illustration of disyllables (like Tone1Tone0, Tone1Tone1, Tone1Tone2, ect.) are that the first syllables of disyllables are on the y axis and the second syllables of disyllable are on the x.

Duoduo							JD						
0;9.27	2nd 1st	T0	T1	T2	T3	T4	0;9.29	2nd 1st	T0	T1	T2	T3	T4
	T1	3	3	0	0	5		T1	1	0	0	0	3
	T2	0	1	0	0	**7**		T2	0	0	0	0	0
	T3	0	0	0	0	0		T3	0	0	0	0	0
	T4	1	1	0	0	0		T4	0	0	0	0	0
0;10.16	2nd 1st	T0	T1	T2	T3	T4	0;10.17	2nd 1st	T0	T1	T2	T3	T4
	T1	1	1	0	0	2		T1	0	0	0	0	4
	T2	0	1	0	0	**5**		T2	0	0	0	0	4
	T3	0	0	0	0	0		T3	0	0	0	0	0
	T4	1	1	0	0	0		T4	0	1	0	0	0
0;11.6	2nd 1st	T0	T1	T2	T3	T4	0;11.13	2nd 1st	T0	T1	T2	T3	T4
	T1	2	3	0	0	5		T1	2	1	0	0	3
	T2	1	2	0	0	**10**		T2	0	0	0	0	0
	T3	0	0	0	0	0		T3	0	0	0	0	0

	T4	3	1	3	0	1		T4	0	1	1	0	0
1;0.3	2nd 1st	T0	T1	T2	T3	T4	1;0.6	2nd 1st	T0	T1	T2	T3	T4
	T1	0	3	0	0	4		T1	0	1	0	0	5
	T2	2	2	0	0	**6**		T2	0	0	0	0	3
	T3	0	1	0	0	0		T3	0	0	0	0	0
	T4	2	3	0	0	1		T4	0	0	0	0	0
Taozi							Wanzi						
0;9.6							0;9.7						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	0	0	0	0	**6**		T1	3	2	0	0	**6**
	T2	0	0	0	0	2		T2	1	0	2	0	6
	T3	0	0	0	0	0		T3	0	0	0	0	0
	T4	1	0	0	0	0		T4	2	0	0	0	0
0;10.3							0;10.4						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	1	1	1	0	**6**		T1	2	3	2	0	**7**
	T2	0	0	0	0	3		T2	2	1	2	0	3
	T3	0	0	0	0	0		T3	0	0	0	0	1
	T4	0	0	0	0	0		T4	2	0	0	0	0
0;11.26							0;11.6						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	4	2	0	0	**8**		T1	5	3	2	0	**8**

	T2	1	0	0	1	7		T2	1	1	1	0	4
	T3	0	1	0	0	0		T3	0	0	0	0	0
	T4	0	0	0	0	0		T4	3	0	1	0	2
1;0.8	2nd 1st	T0	T1	T2	T3	T4	1;0.29	2nd 1st	T0	T1	T2	T3	T4
	T1	5	3	1	0	**15**		T1	6	3	0	0	**7**
	T2	0	2	0	0	6		T2	0	0	0	0	2
	T3	0	0	0	0	0		T3	0	0	0	0	1
	T4	0	1	0	0	0		T4	4	0	2	0	2
Shiliu							Kele						
0;9.26							0;9.7						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	3	**9**	3	0	4		T1	0	**6**	1	0	3
	T2	3	1	0	1	5		T2	2	1	0	0	**9**
	T3	0	0	0	0	0		T3	0	0	0	0	0
	T4	3	0	1	0	1		T4	2	2	2	0	0
0;10.22							0;10.12						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	5	**7**	2	0	3		T1	1	**7**	0	0	1
	T2	1	4	1	0	3		T2	1	1	1	1	**7**
	T3	0	0	0	0	0		T3	0	1	0	0	1
	T4	2	0	2	0	0		T4	2	1	2	0	1

0;11.12

2nd 1st	T0	T1	T2	T3	T4
T1	4	**7**	3	0	5
T2	0	3	1	0	5
T3	1	0	0	0	0
T4	1	0	1	0	2

0;11.9

2nd 1st	T0	T1	T2	T3	T4
T1	0	**10**	3	0	5
T2	3	0	0	0	**7**
T3	0	0	1	0	2
T4	2	2	0	0	2

1;0.13

2nd 1st	T0	T1	T2	T3	T4
T1	5	**8**	1	0	4
T2	0	3	1	0	3
T3	0	0	0	0	0
T4	2	1	0	0	2

1;0.14

2nd 1st	T0	T1	T2	T3	T4
T1	1	**6**	1	0	4
T2	0	1	2	0	**8**
T3	0	0	0	0	0
T4	2	3	0	0	0

Lele

0;9.21

2nd 1st	T0	T1	T2	T3	T4
T1	2	3	0	0	**7**
T2	2	0	1	0	**7**
T3	0	0	0	0	0
T4	0	0	1	0	0

0;10.20

2nd 1st	T0	T1	T2	T3	T4
T1	2	9	5	0	**13**
T2	3	3	1	0	**16**

Wenwen

0;9.7

2nd 1st	T0	T1	T2	T3	T4
T1	1	2	1	0	1
T2	2	0	0	0	4
T3	0	0	0	0	0
T4	2	0	0	0	1

0;10.23

2nd 1st	T0	T1	T2	T3	T4
T1	0	1	0	0	2
T2	0	0	0	1	3

	T3	0	0	0	0	1		T3	0	0	0	0	0
	T4	1	7	4	0	4		T4	0	0	1	0	2
0;11.6							0;11.2						
	2nd 1st	T0	T1	T2	T3	T4		2nd 1st	T0	T1	T2	T3	T4
	T1	1	3	1	0	**8**		T1	0	1	2	0	2
	T2	1	2	1	0	**9**		T2	0	0	0	0	1
	T3	0	0	0	0	0		T3	0	0	0	0	0
	T4	1	0	1	0	1		T4	1	0	1	0	1
1;0.13	2nd 1st	T0	T1	T2	T3	T4	1;0.9	2nd 1st	T0	T1	T2	T3	T4
	T1	1	4	1	0	**9**		T1	3	2	0	0	4
	T2	1	0	0	0	**7**		T2	2	1	1	1	5
	T3	0	0	1	0	1		T3	0	0	0	0	0
	T4	1	0	0	0	1		T4	3	2	0	1	0

Qiqi

0;9.14						
	2nd 1st	T0	T1	T2	T3	T4
	T1	1	1	0	0	**6**
	T2	0	0	1	0	1
	T3	0	0	0	0	0
	T4	1	0	1	0	0
0;10.26						
	2nd 1st	T0	T1	T2	T3	T4

	T1	1	5	1	0	**6**
	T2	0	0	1	0	2
	T3	0	0	0	0	0
	T4	1	0	0	0	0
0;11.18						
	2nd 1st	T0	T1	T2	T3	T4
	T1	0	5	1	0	**7**
	T2	0	0	0	0	1
	T3	0	0	0	0	0
	T4	0	0	0	0	0
1;0.9	2nd 1st	T0	T1	T2	T3	T4
	T1	1	6	1	0	**7**
	T2	0	0	0	0	3
	T3	0	0	1	0	0
	T4	1	1	0	0	0

Appendix II. Occurrence of tones in York English-learning children's disyllabic utterances (tone combinations that are not produced by any child are not listed; bold face indicates the most frequent pattern in the given session)

Name	Age	T1-T1	T2-T2	T2-T4	T4-T1	T4-T2	T4-T4	Total	Uncat.
AC	0;9.		3	**13**		4	12		1
	1;0.		1	**18**		7	11		5
AH	0;9.		1	**6**		3	**6**		3
	1;0.		1	10			**12**		
CL	0;9.			**10**		5	**10**		3
	1;0.			**8**		4	2		
CS	0;9.	1		3			**5**		
	1;0.	1		3			**5**		
GR	0;9.	1		10	2	1	**16**		
	1;0.			1			**3**		
JY	0;9.			2	2	2	**3**		
	1;0.		1	**6**	1	**6**	**6**		
RT	0;9.			**16**		2	11		1
	1;0.		1	**12**		4	10		1
S	0;9.		1	**8**		1	6		1
	1;0.		2	10		7	**22**		5

Appendix III. Mandarin children's use of consonants and vowels at 4wp

	Keke	Xinyu	Yifan	Yiyi	Zuozuo	Didi	Shi	PingAn
p	✓	✓	✓	✓	✓	✓	✓	
t/t	✓		✓	✓		✓	✓	✓
k	✓		✓			✓		
m		✓		✓	✓	✓	✓	✓
n	✓					✓	✓	
ɕ				✓				
tɕ			✓			✓		
l								✓
j			✓					
w		✓			✓		✓	
i		✓	✓	✓				
e	✓	✓	✓			✓		
ɛ	✓	✓	✓			✓	✓	✓
a	✓	✓	✓	✓	✓	✓	✓	✓
ə						✓		
u	✓				✓	✓	✓	✓
o		✓			✓			
ɤ	✓							
au	✓				✓			

ou	✓				
ai			✓		
ei		✓			
iu	✓				
iɛ			✓	✓	✓
ia				✓	

Appendix IVa. Zuozuo's words at 15 months, ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 14 (.61)			
妈妈 ma1ma0	ma1, ma2(2)	没 mei2	ma2
爸爸 pa4pa0	pa4	给 kei3	ka4
不 pu4	pu4	在 tsai4	ka4
拔 pa2	pa2	喂 wai2	wa2(8), wa3(2)
ball	pa4	饼干 piŋ3kɛn1	pa2(2)
打(电话)ta3	ta2(2), ta3, ta2-4	拜拜 pai1pai2	pai3, pai2(2)
这 ʈʂɤ4	tɛ2	摸 muɤ1	ma1
CVV 4 (.17)			
花 xua1	xua1		

拍 p^hai1	pai4
消失又出现 mau2	mau2, mou3
C_1VC_1V 2 (.09)	
爸爸 pa4pa0	pa1pa0(2)
妈妈 ma1ma0	ma1ma0(3), ma2ma1, ma2ma0(9), ma2ma2
C_1VC_2V 1 (.04)	
不动 pu2tʊŋ4	pu1tɤ4
single C & glide/glottal 1 (.04)	
吃饭声音 a4mu0	a4mu0
$CVCVCVV_0$ 1 (.04)	
拔不动 pa2pu0tʊŋ4	pa2pu0tɤ4, pa2pu0tio4

Appendix IVb. Yifan's words at 16 months, ordered by prosodic structures.

selected words		adapted words	
target	child form	target	child form
CV 5 (.)			
怕 pa4	pa4	妈妈 ma1ma0	ma1(5)
不 pu4	pu1	嘎嘎 ka1ka1	ka4(4)
起 $tɕ^hi3$	$tɕ^hi4$		

CVV 1 (.04)

拜拜 pai1pai2	pai2

C_1VC_1V; CVV_0CVV 11 (.)

娃娃 wa2wa0	wa4wa4	这 ʈʂɛ4	ʈʂɛ1ʈʂɛ1
妈妈 ma1ma0	ma1ma0	喵喵 miao1miao1	mɛ1mɛ1
嘎嘎 ka1ka1	ka4ka0, ka1ka4, ka4ka4(2), kua1ka2, ka3ka4	姐姐 tɕiɛ3tɕiɛ0	dɛ1dɛ4(2), dɛ4dɛ0(2) tɕiɛ1tɕiɛ4(3)
爸爸 pa4pa0	pa4pa4		
打打 ta2ta0	tɛ1tɛ4, tɛ2tɛ4, ta1ta2		
鸽子 kɤ1tsi3	kɤ1kɤ2, kɤ4kɤ4		
折纸 ʈʂɤ2ʈʂi3	ʈʂa2ʈʂe3		
琴键 tɕʰin2tɕiɛn4	tɕia4tɕia0, tɕi4tɕia4		

C_1VC_2VV, C_1VVC_2V 3(.)

不乖 pu4kuai1	pu4kua4	这边 ʈʂɛ4piɛn1	tɛ1tʰuɚ2, tiɛ4pu0, tɛ1pʰu4,
就这 tɕiu4ʈʂɤ4	tɕiu4ʈʂə0		

single C & glide/glottal 2 (.)

哎呀 ai1ja0	ai1ja0, ei1ya4
二姐 ɚ4tɕiɛ3	xɛ4tɕiɛ2, ɛ4dɛ4