Chapter 2 Literature review

2.1 Phonological Development in English-speaking children

In this section, I will first describe the stages of sound production for early speaking children. I will then present the acquisition studies of how children learn consonants.

The development of the speech production has been viewed as a sequence of stages. The order of the stages appears to have cross-linguistic aspects in children's language acquisition. The order of the stages in the speech production before infants begin to produce their "first word" is given according to Stark's proposal (1980) as shown below.

Stage1. Reflexive vocalization	0-6 weeks	These vocalizations include cry and fussing sounds and vegetative sounds which do not exist later.
Stage 2. Cooing and laughter	6-16 weeks	These sounds are drawn out by the interaction with adults or older kids. But, later infants may produce these sounds without the interaction. However, smiling, or nodding faces are still powerful stimuli to the infants.
Stage 3. Vocal play	16-30 weeks	Sounds produced in this stage have the characteristics of playful use of behavior such as squealing, growling, yelling loudly, production of noise by blowing air and food.
Stage 4. Reduplicated babbling	31-50 weeks	A series of consonant-vowel (CV) syllables start to appear in this stage.
Stage 5. Nonreduplicated babbling	10-14 months	In this stage, in addition to consonant-vowel syllable, the vowel (V), vowel-consonant (VC) and consonant-vowel-consonant (CVC) syllables come into infant's production. Besides, a variety of stress pattern and intonation tone is found in the later stage of babbling.
Stage 6. Single word production	period of variable duration	The production of infants includes (a) protowords and (b) words used as symbols and refer to recurring objects or events.

In the stages of speech production, we may see each stage has its own typical behavior. The typical behavior may have precursors in the previous stages and continue into the following stages (Stark 1980). In Stage 4 (31-50 weeks), the syllable (consonant-vowel) has become a unit in the speech production. The production of CV may have the precursors in the previous stages, such as cooing, laughter, and vocal play. The production of CV may continue into stage 5 (Stark 1980).

In the acquisition of the initial consonants, the English data discussed by Macken (1980) have presented that voiceless unaspirated stops are acquired earlier than the patterns of voicing types. There is a high probability that children acquire the front stops prior to the back stops. In the data, Macken found that children would substitute the dental stops for the velar ones. As to the acquisition of the initial fricatives and affricates, they are considered to be the most difficult parts for English-speaking children (Ingram, Christensen, Veach & Webster. 1980). First, those sounds are kept away from very young children's usage (Ferguson, 1973). Second, they are the last sounds to be acquired by normal older children (Templin 1957). They are also the most difficult sounds for children with articulatory delay (Morley 1957). Despite of the difficulties of fricatives and affricates for children in the acquisition, the data of Ingram et al. (1980) have shown a general pattern that [❄-] before [✳-, ▼✳-, ❄✺-] and they are before [▲-, ❖-, ▮-, ✶-]. However, not all the children follow the general patterns in their acquisition. The individual variations are quite considerable in terms of the sequence of the acquisition.

2.2 Phonological development in Mandarin speaking children

2.2.1 The Sketch in Mandarin Phonology

The description of the inventory in Mandarin Phonology is based on the studies from Cheng (1973), Lin (1989), and Wan (1999).

2.2.1.1 Consonants

Mandarin in Taiwan has its own phonetic symbols, *Mandarin phonetics alphabet Zhuinfuhao*. However, throughout this study, the Mandarin phonetic phonemes are presented by the IPA system. The possible consonants in Mandarin are shown in Table 1.

Table 1 : Mandarin consonants (e.g. Wan 1999:17)

	Bilabial	**Labial**	**Dental**	**Retroflex**	**Palatal**	**Velar**
Unaspirated						
Plosive	p		t			k
Aspirated						
Plosive	pʰ		tʰ			kʰ
Fricative		f	s	ʂ/ʐ	ɕ	x
Unaspirated						
Affricate			ts	tʂ	tɕ	
Aspirated						
Affricate			tsʰ	tʂʰ	tɕʰ	
Nasal	m		n			ŋ
Liquid			l			
Glide					(j)(ɥ)	(w)

In Table 1, there are 24 possible consonant phones in the inventory of Mandarin phonology. The places of the articulation are composed of 6 parts (bilabial, labial, dental, retroflex, palatal and velar) and the manners of the articulation are composed of 8 kinds (unaspirated, plosives, aspirated, fricative, affricate, nasal, liquid and glide).

2.2.1.2 Vowels

The vowel inventory is drawn from Cheng (1973), Lin (1989) and Wan (1999) and is shown in Table 2.

Table 2: Mandarin Vowels (Wan 1999:24)

	Front	Central	Back
Close	i y	ɨ	u
Close-mid	e		ɤ o
		ə	
Open-mid	ɛ		ɔ
Open	a		ɑ

In Table 2, there are 12 possible vowel phones in the inventory. The vowels are defined in terms of the places (front, central and back) and the mouth shapes from close to open (close, close-mid, open-mid and open).

2.2.1.3 Syllables

Mandarin is considered to have the following syllable structures: **V, CV, GV, VG, VN, CVN, GVN, CGVG** and **CGVN** (V= vowel, C= consonant, G= glide and N= nasal) as shown in Table 3.

Table 3: Possible syllable in Mandarin (Wan 1999: 36)

Syllable structure	**Phonetic transcription**	**Gloss**
V	i 55	dependent
CV	mɑ 55	mother
GV	ja 55	push
VG	aj 51	love
VN	an 55	safe
CVG	maj 214	buy
CVN	tiŋ 214	top
GVG	jɑw 35	shake
GVN	wan 51	ten thousands
CCVG	tiɑw 51	drop
CGVN	tjɛn 55	east

In the acquisition of consonants, it has been found that aspiration feature is acquired earlier than the retroflex (Li, 1978; Hsu, 1987, Shiu 1990). Hsu (1987) has presented a division of stages in the acquisition of initial consonants and the division of stages listed in Table 4.

Table 4: the stages in the acquisition of Mandarin initial consonants (Hsu 1987)

Stage 1	1;0-1;8	Affricates and fricatives are often replaced by stops.
Stage 2	1;8-2;6	Extensive emergence of aspirated stops is observed.
Stage 3	2;7-3;2	Emergence of fricatives and affricates except for the retroflexes
Stage 4	3;3-4;3	Emergence of retroflexes, but not stable
Stage 5	4;4-6;0	completion of acquisition

Besides, a cross-sectional study by Wang, Fei, Huang and Chen (1984) also reveals that the consonants [❐, ❐★, ❍, ▼, ▼★, ■, ●, ✳, l] are acquired at the age of 3 and [▼⇛, ▼❧, ❧] and [➑] are acquired after the age of 6.

With regard to the acquisition of vowels, Hsu (1987) found that [ɑ] emerges first and then is followed by [i], which precedes [u] in the period from 1;0 to 1;2. The primary vowels such as [i, ɑ, u, ɔ, ✣, ə] are observed to appear earlier than the glides [j, w] (Su, 1985). The appearance of [j] is also late in the study by Shiu (1990). As to the diphthongs, Hsu (1987) found that although they emerge almost as early as the single vowels, 6-years-old children still have problems with the off-gliding diphthongs [ei] and [ou]. Besides, Li (1978) reported that off-gliding diphthongs are often monophthongized.

2.3 The development of infants' speech perception

Perception develops earlier than production. Before infants are born, they can actually hear (Gleanson & Ratner 1998). Soon after they are born, they show the preference of their mothers' voices (Gleanson & Ratner 1998). Infants present certain behaviors when they perceive a novel or new sound. Two methods, HASP (high-amplitude sucking paradigm) and cardiac deceleration, are used to exam children's responses to the stimuli. HASP (Eimas, Jusczyk & Vigorito 1971) is a method which detects infants' rate of pacifier-sucking. The experimenters let infants listen to a new sound (for example, [ba]). As the new sound comes out, infants speed up the sucking rate. While the new sound [ba] has continuously been produced, the sucking rate will get back to the normal speed. Then, the experimenter lets the infants listen to another new sound (for example, [pa]). If infants' sucking rate remains the same, that means they are not able to contrast the two sounds, [ba] and [pa]. If the sucking rate increases, it indicates that they can discriminate the two sounds, /ba/ and /pa/. The cardiac deceleration is the method which detects infants' rate of hear-beating. It is much the same way as HASP. Infants' rate of hear-beating increases when they are provided with new stimuli. It has been found that infants are able to discriminate the input utterances which are different by one feature in the onset position (Eimas et al., 1971). For example, they are able to distinguish the onset phonemes /p/ and /b/ in English. Furthermore, infants from one month to six months have shown the ability to contrast the features of the speech sounds which are not used in their environment (Gleason & Ratner 1998). For instance, Werker & Tees (1984) found that Canadian babies can contrast Czech /ř/ from /ʒ/, although the language Czech is not used in babies' environment. However, this ability for contrasting the features of foreign language gradually loses because this ability has no space to develop or is not required in infants' linguistic environment (Studdert-Kennedy 1987).

Compared with the methods used in infants, methods used for children or adults have made two kinds: discrimination and identification (Gleasin & Ratner 1998). Discrimination is the method that requires the listeners to judge whether the stimuli are the same or different. For example, the listeners are asked to judge whether the sounds [ba] and [pa] are the same or not. Identification is the method that requires the listeners to make decision on the status of the stimuli. Besides, another design for perception is the classification task. Classification is the method that requires listeners to categorize words according to an occurring signal. In the classification tasks (Treiman & Baron 1981, Treiman & Breaux 1982), children aged 5 years are asked to judge and classify whether the sounds are similar or not. For examples, they will be asked whether /bɛz/ & /vis/ are similar, or /bɛz/ & /bug/. Compared with /bɛz/ & /bug/, which share one phoneme at the beginning, /bɛz/ & /vis/ have no correspondence in any single phoneme. However, the results show that children tend to judge /bɛz/ & /vis/ as similar sounds instead of /bɛz/ & /bug/. One explanation is that /bɛz/ & /vis/ are similar overall across the whole sound (Singh, Woods & Becker 1972). Contrast to the results by adults, adults tend to classify /bɛz/ & /bug/ together because the initial phonemic correspondence is important for them to make classification. The results have shown that young children have the tendency to focus on the overall shape of the sounds in the judgments and classification of the perceived sounds.

2.4 Early phonological representation

From the study of infant speech perception (Jusczyk, 1986), the function of the representation for the speech sounds is to differentiate the utterances spoken by adults. In 1970s, there is an intensive work on child phonology to investigate what is presented in children's mental lexicon. In order to describe the "representation" in children's mental lexicon, the description used in adult language, phonemes and syllables, was adopted to describe the representation of the words in children's mind (Studdert-Kennedy, 1987).

Children's phonological performances have something to do with the phonological representation. As for the lexical representation, the study from Aslin

& Smith (1988) shows that young children' s representation of lexicon is holistic in nature. For example, in the early phonological process, the representation of /dɔg/ is not organized as sequential segments like /d/ + /ɔ/ + /g/. Instead, words are represented on the basis of individual salient characteristics of overall acoustic shape (Waterson 1971).

As the amount of vocabulary grows, there is an increased pressure to make the representation of words more specific in order to differentiate them (Jusczyk 1987, Walley 1993). Under the pressure of the explosive vocabulary growth, there is a need to discriminate the speech sounds efficiently in real time production or perception. As the result, the rapid growth of vocabulary leads to the developmental change in the word representation in which the word representation is more specific (Walley 1993). In order to achieve adult-like levels, children have to refine the internal organization (or the representation) of speech sounds according to the characteristics of their language. For examples, for English speaking children, they need to modify the representation of /kæt/ into /k/+ /æt/ when they encounter words like /bæt/ in which the initial phoneme is different and the rest of the parts is the same. Therefore, in the overview of phonological development, children's representation of speech sounds has changed in the trend from holistic to segments undergoes a change from a more holistic level to a more segmental level.

2.5 Effect of the size of a unit

The size of a unit has been found to be a primary influence on children's performances. The size which is measured by the numbers of phonemes may affect children's performances in some kinds of sound classification tasks (Walley, Smith & Jusczyk 1986). For example, when children hear the sound pairs [bɔrn]-[bʌmp], [zɪbrɑ]-[zɪple], [sɑlɪd]-[sɑləm] and [kutɚ]-[dɪle], they have to judge whether those stimulus pairs have shared any correspondent units. The correct answer for [bɔrn]-[bʌmp], [zɪbrɑ]-[zɪple] and [sɑlɪd]-[sɑləm] would be "yes" because those pairs have shared the same units. The word-pairs [bɔrn]-[bʌmp] share the onset [b] , [zɪbrɑ]-[zɪple] share the syllable [zɪ] and [sɑlɪd]-[sɑləm] share the three phonemes [sɑl]. With regard to the size, [bɔrn]-[bʌmp] share the least phonemes (one phoneme) correspondence, and [sɑlɪd]-[sɑləm] share the maximal (three phonemes) correspondences. The correct answer for [kutɚ]-[dɪle] would be "no" because they don't share any of the same phonemes.

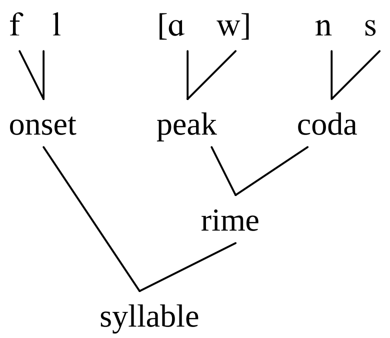
The effect of size refers to improved performances on sound classification as the size of the shared units increases. Children did better in syllable condition than in onset/rime condition, which in turn is better than the single phoneme condition (Treiman & Zukowski 1991). Besides, the significant correlation between the numbers of the shared units and phonological awareness scores (the performance) has been observed in the study of Brady, Fowler & Gipstein (1994). In addition, Walley et al. (1986) suggest that kindergarteners like to compare speech sounds through lager units than syllabic units. They found that the kindergarteners performed better on the type with the maximal shared units (three phonemes correspondence) in the tasks of sound classification.

2.6 Theoretical views on the syllabic structures

The importance of syllables is not only observed in children's performances on the phonological tasks but also in the theories of syllables. Selkirk (1982), who takes English as an example, provides three points in support of the syllable as a theoretical construct. First, the phonotactic constraints can only be made by the reference to the syllabic structures. Second, the syllable gives a proper domain for the application of the phonological rules. Third, a syllable also serves as an ideal bearing unit for the suprasegmental phenomena such as stress and tones. In addition, Blevins (1995) has offered the arguments on the importance of syllables in the phonological theory in term of the following perspectives: syllable as a domain, syllable edge as locus, syllable as target structures, and native intuition. As to syllable as a domain, Blevins (1995) points out the phonological properties such that tone and stress take syllables as their domain. As to syllable edge as locus, aspiration is often associated with syllable boundaries. For instance, in English the syllable initial-obstruents are aspirated. As to syllable as target structures, syllable can function as the targets of language game. With the survey of language games, the syllable is often the target of suffixation, truncation, substitution, or movement in at least twenty cases (Laycock 1972). As to native intuition, many native speakers have clear intuitions with regard to the numbers of the syllables in their languages or take the syllable as a unit to break the strings of sounds (Blevins 1995).

With regard to the internal phonological representation of an English monosyllabic word, the word “flounce” is structured as in (1) (Selkirk 1982).

(1)



Looking at the phonological representation in (1), the monosyllabic English word is organized with an onset [fl] and a rime which is structured by a peak [ɑw] and a coda [ns]. With regard to the movement or the application of phonological rules on the sub-constituents of the syllables, the operations on the components of the syllables are dependent and restricted (Sekirk 1982). For example, in Mandarin the onset [f] can combine with the rimes such as [ɑ], [u] and [ej], but not with the rime such as [ɑw] and [i] .

According to the theoretical views on the syllable and its sub-constituents, the syllable is suggested to serve as an optimal unit for the phonological processing whereas the phonological operations on the sub-constituents (phonemes) are more complicated regarding to the phonotactic constraints. In this view, the syllable may have the advantages over its intrasyllabic units because of the restrictions to the sub-constituents.