

ARTICLE

# Development of stop consonants in three- to six-year-old Mandarin-speaking children

Jing YANG\*†

Department of Communication Sciences and Disorders, University of Central Arkansas

\*Corresponding author. E-mail: [jyang@uca.edu](mailto:jyang@uca.edu)

†Jing Yang, Department of Communication Sciences and Disorders, University of Central Arkansas, 201 Donaghey Ave, Conway AR 72035. tel: 501-450-5486.

(Received 15 May 2017; revised 30 September 2017; accepted 14 March 2018;  
first published online 18 April 2018)

## Abstract

This study compared the temporal measurements of stop consonants in 29 three- to six-year-old Mandarin-speaking children and 12 Mandarin-speaking adults. Each participant produced 18 Mandarin disyllabic words which contained six stop consonants /p, p<sup>h</sup>, t, t<sup>h</sup>, k, k<sup>h</sup>/ each followed by three vowels /a, i, u/ at the word-initial position in the first syllable. The temporal measurements of VOT, overall burst duration, average duration per burst, number of bursts, and VOT-lag duration were obtained. Although adult-like short-lag VOTs were achieved in all children, the long-lag VOTs were widespread in the younger group and gradually developed to a concentrated distribution in the older children. Further analysis of the burst and VOT-lag revealed that these children tended to produce shorter average duration per burst and longer VOT-lag than the adults. These results indicate that children in this age range may not have developed adult-like laryngeal–oral timing pattern and airflow control for stop production.

**Keywords:** stop consonants; Mandarin-speaking children; durational features

Research on phonological acquisition of speech segments has shown that, cross-linguistically, stop consonants are always acquired earlier than the other types of consonants and are normally acquired before four years of age (Dodd, Holm, Hua, & Crosbie, 2003; Hua & Dodd, 2000; Smit, Hand, Freilinger, Bernthal, & Bird, 1990; So & Dodd, 1995; Templin, 1957; Wellman, Case, Mengert, & Bradbury, 1931). Substantial previous research has examined the development of stop consonants, in particular the voicing feature represented by the measure of Voice Onset Time (VOT), in English-speaking children (Hitchcock & Koenig, 2013; Kewley-Port & Preston, 1974; Macken & Barton, 1980a; Zlatin & Koenigsknecht, 1976). Sporadic reports have also been published on the acquisition of contrastive production of stop consonants in children from other language backgrounds (e.g., Davis, 1995; Gandour, Petty, Dardarananda, Dechongkit, & Mukngo, 1986; Lee & Iverson, 2008; Macken & Barton, 1980b). Extending from previous studies, the present investigation aims to document and examine the temporal features of stop consonants in

Mandarin-speaking children aged three to six years old. Acoustic measures including VOT, overall burst duration, average duration per burst, number of bursts, and duration of VOT-lag will be examined and compared in Mandarin-speaking children and adults to provide a detailed profile of the acoustic development of stop consonants in this population.

### *VOT development in English-speaking children*

For stop sounds occurring in word-initial position, the temporal organization of the release of oral occlusion, the expiration of airflow, and the vocal fold adduction characterizes the voicing and aspiration features of the stop consonants. VOT, defined as the time interval between the release of stop closure and the onset of vocal fold vibration, has been widely acknowledged as the primary cue for identifying the voicing feature and for differentiating categories of word-initial stops across different languages all over the world. According to early cross-language studies (Abramson, 1977; Lisker & Abramson, 1964), stop consonants can be divided into three general categories along the VOT continuum: voicing lead (<0 ms), short voicing lag (0–20 ms), and long voicing lag (>40 ms). Voicing lead corresponds to fully voiced stops for which the onset of voicing occurs before the release of oral occlusion. Short-lag VOT represents voiceless unaspirated stops for which the onset of voicing occurs shortly after the oral release. Long-lag VOT characterizes voiceless aspirated stops for which the oral release is immediately followed by an expiration of airflow before the start of phonation. During the process of acquisition of stop consonants, children need to learn how to precisely place the articulators, adopt the appropriate aerodynamic mechanism, and form the articulatory sequence for stops with different places of articulation. Furthermore, they need to establish the language-specific VOT pattern to accurately present the voicing and aspiration features of their native language. As VOT involves the complex articulatory coordination of laryngeal and supralaryngeal musculature, this acoustic measure has also long been used to signify the maturation of speech timing control from children to adults (Koenig, 2000; Nittrouer, 1993).

English has six stops produced at the bilabial, alveolar, and velar regions. The six stops represent a voiced–voiceless distinction. However, the VOT values of English stops do not strictly conform to the VOT categorizations for the voiced–voiceless distinction (Lisker, 1986). English voiceless stops /p, t, k/ in syllable-initial position are normally produced with aspiration, with the average VOTs located in the long-lag range. English voiced stops /b, d, g/ are commonly produced with absence of vocal fold vibration, and are presented as short-lag VOTs, although some speakers in some dialects produced /b, d, g/ with VOT in the voicing-lead category (Keating, 1984; Klatt, 1975; Lisker, 1986; Lisker & Abramson, 1964).

Numerous studies have been conducted to address the acquisition of phonological and phonetic contrasts of stop consonants in English-speaking children (e.g., Bernthal, Bankson, & Flipsen, 2009; Bond & Wilson, 1980; Engstrand & Williams, 1996; Gilbert, 1977; Hitchcock & Koenig, 2015; Kewley-Port & Preston, 1974; Lowenstein & Nittrouer, 2008; Macken & Barton, 1980a; Preston & Yeni-Komshian, 1967; Preston, Yeni-Komshian, Stark, & Port, 1968; Whiteside & Marshall, 2001; Zlatin & Koenigsknecht, 1976). Previous researchers have claimed that the English voicing contrast was fully acquired by age three, and have summarized a general three-stage developmental path of VOT in English-speaking children. During the

initial development course, VOT values demonstrated a unimodal pattern with most stops produced with short-lag VOTs. Then, children started to produce intended voiceless stops with longer VOTs than the intended voiced stops, but the long-lag VOTs for voiceless stops were not as long as the adult targets, which caused an emerging bimodal VOT pattern, but short-lag VOTs were still the primary model. In the third stage, the long-lag VOTs refined to approximate the adult values. Some children experienced an ‘overshoot’ phase with extremely long long-lag VOTs. During this stage, a typical bimodal pattern with clearly separated short-lag and long-lag VOTs was developed (e.g., Gilbert, 1977; Lowenstein & Nittrouer, 2008; Macken & Barton, 1980a; Preston *et al.*, 1968; Zlatin & Koenigsnecht, 1976).

Among different places of articulation in stop consonants, the dichotomy of voicing contrast emerged later in velar stops than in alveolar or bilabial stops (Barton & Macken, 1980). Compared to adult norms, children generally demonstrated greater variability and more dispersed VOT distribution. The widely dispersed VOT causes overlap between phonetic categories in children (Imbrie, 2005; Koenig, 2000, 2001, Romeo, Hazan, & Pettinato, 2013). According to Koenig (2000), the greater variability in children involves laryngeal factors such as abduction degree and vocal fold tension, in addition to the developing inter-articulator timing control. Researchers proposed supplementary measures of range, accuracy, discreteness, and overshoot of VOT values to better quantify the distributional features of VOT data in children younger than two years of age (Hitchcock & Koenig, 2013; Koenig, 2001). As children grew older and became more mature in articulatory control, the VOT values and the dispersion of VOT distribution decreased (Lee & Iverson, 2012; Romeo *et al.*, 2013).

### *Durational and burst features beyond VOT*

Speech production involves three processes: respiration, phonation, and articulation. The production of stop consonants can normally be described by three phases: closing, holding, and releasing. During the closing phase, a complete constriction is formed at a certain location along the vocal tract, and the velopharyngeal port closes to avoid air escaping from the nasal cavity. The speaker holds the closure for a short period to build up the intraoral pressure, during which the status of the glottis and vocal folds are adjusted to characterize the voicing feature for voiced or voiceless stops. When the intraoral pressure is built up to a certain point, the oral closure is released, which generates a transient burst. This articulatory sequence requires precise coordination between oral and laryngeal muscular movements. It is reflected in acoustic representations and phonetic correlates such as the duration, amplitude, and spectral prominence of the burst and aspiration, formant transition to neighbor vowels, and F0 at the voicing onset, in addition to the most commonly used measurement of VOT. Among these measures, the temporal features are of particular interest in the present study because they are related to all three processes of stop production (Imbrie, 2005, Table 1 on p. 33).

Regarding the respiratory process, a relatively high and strong airflow likely pushes the oral constriction more quickly and thus reduces the burst duration. For phonation, the voicing feature and degree of aspiration are reflected in the durational pattern of VOT. As to the articulatory process, according to the aerodynamic explanation (Stevens, 2000), the area of articulatory contact at the constriction, the length of constriction, and the rate of cross-sectional area change determine the amount of

**Table 1.** VOT Means of Stop Consonants in Mandarin (M) or Taiwanese Mandarin (TM) Reported in Previous Studies. Note that the Data in Li (2013) Were Reported Separately for Females and Males.

	Rochet & Fei (1991) (M)	Chen <i>et al.</i> (2007) (TM)	Chao & Chen (2008) (TM)	Li (2013) (M)
p		13.9	14	
t		15.3	16	12.5 (F), 17.5 (M)
k		27.4	27	22.5 (F), 29.5 (M)
p <sup>h</sup>	99.6	77.8	82	
t <sup>h</sup>	98.7	75.5	81	93.3 (F), 77.9 (M)
k <sup>h</sup>	110.3	85.7	92	90.8 (F), 78.8 (M)

time elapsed on stop closure and release burst. In particular, the closure duration varies with the degree of intraoral pressure behind the constriction. Velar stops have a smaller cavity behind the constriction, which requires less time to build up the intraoral pressure in comparison to bilabial stops. Therefore, velar stops normally have a shorter stop closure than bilabial stops. Meanwhile, as velar stops have the constriction at a position further back, formed with a large area of articulatory contact between the back of the tongue and the soft palate, the change of cross-sectional area and intraoral pressure for velar stops is relatively slow in comparison to that in bilabial and alveolar stops. As a result, velar stops are normally produced with multiple bursts and longer burst duration than the other two.

The articulatory mechanism and its effect on the durational features of stop components have been examined in adult speakers (Maddieson, 1997; Stevens, 2000). However, the extent to which physiological and aerodynamic factors affect the acoustic characteristics of stops in children has rarely been investigated. So far, one study by Imbrie (2005) has conducted an in-depth acoustical analysis for stop consonants in English-speaking children aged between 2;6 and 3;3. As for the durational and burst measurement, Imbrie compared VOT, overall burst duration, average burst duration for single burst, average duration per burst, number of bursts, VOT-lag duration, etc. Imbrie found that while the children did not show a significant difference from the adults on the average overall burst duration, they produced significantly shorter average burst duration for single burst and average duration per burst with a high prevalence of multiple bursts. A possible explanation for this result was that children had a smaller articulator size and initiated high subglottal pressure at the beginning of speech utterances, which released the constriction more quickly and reduced the burst duration in comparison to adult speakers. Meanwhile, the high intraoral pressure was typically followed by negative pressure which caused the constriction again for subsequent bursts. Therefore, children tended to produce multiple bursts in comparison to adults. In addition, Imbrie found that English-speaking children had a significantly delayed VOT that was represented by the measure of VOT-lag duration for voiced stops, which suggested that children at this age had not developed adult-like laryngeal coordination between the relaxing of the vocal folds and the stop release. For the voiceless stops, Imbrie found significantly longer VOT in children than in adults, and the VOT values displayed a decreasing pattern with age.

### Present study

Similar to English stops, Mandarin Chinese also contains six stop consonants /p, t, k, p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ produced in bilabial, alveolar, and velar positions. All Mandarin stops are voiceless and only occur in word-initial position. The six stops represent an unaspirated–aspirated contrast. Previous studies have reported VOT data in adult speakers of Mandarin or Taiwanese Mandarin (Chao & Chen, 2008; Chen, Chao, & Peng, 2007; Li, 2013; Rochet & Fei, 1991) (see Table 1 for details). Although these studies differed in specific VOT values, the measurements showed that Mandarin stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ have long VOT values greater than 75 ms, and that /p, t, k/ are produced with short VOT values in the range of 10–30 ms. Following the three-category model proposed by Lisker and Abramson (1964), Mandarin unaspirated stops /p, t, k/ occupy the short-lag regions, while Mandarin aspirated stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ occupy the long-lag regions.

Given that syllable-initial stops in English are commonly produced with short-lag and long-lag VOTs, Mandarin and English stops have similar VOT patterns, with VOT values falling in similar regions along the continuum. However, it is noteworthy that Mandarin stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ are typically produced with strong aspiration, while English syllable-initial /p, t, k/ are produced with weaker aspiration. The VOTs of Mandarin aspirated stops are usually longer than those of English voiceless stops (Chen *et al.*, 2007).

Since Mandarin and English stops are produced at similar places with similar phonetic representations of VOT, the question of interest is whether Mandarin-speaking children show compatible developmental patterns of stop consonants with English-speaking children. In the present study, we addressed this question from two perspectives: first, as VOT has been widely used as the dominant measure for stops and there have been a large number of studies examining VOT development in English-speaking children, the present study examined the development of short-lag and long-lag VOTs for the Mandarin unaspirated–aspirated distinction in Mandarin-speaking children. Second, as shown in the abovementioned study of Imbrie (2005) for English-speaking children, the acoustic development of stops was not only reflected in the VOT values but also manifested in other detailed durational and burst features associated with the physiological constraints and continuing maturation of articulatory control. Following this study, we investigated the development of durational and burst features in addition to the VOT values in Mandarin-speaking children. Additionally, as the detailed acoustic features of stop components have not yet been examined in the Mandarin-speaking population, but Mandarin and English show great similarity on the places and VOT representations of stop consonants, we wonder whether these features of Mandarin stops are affected by the articulatory constraints (e.g., place of articulation and vowel context) in a similar way to English stops.

Previous phonological studies suggest that Mandarin-speaking children and Taiwanese Mandarin-speaking children acquired stops by age 3;0 (Hua & Dodd, 2000; So & Zhou, 2000). However, no acoustic study has been published on stop development in Mandarin-speaking children. One longitudinal study by Pan (1995) documented the development of stop VOT over a five-month period in two Taiwanese-speaking girls at 28 and 29 months old, respectively, which could be of relevance to the present study. Taiwanese, based on the Southern Min dialect of Chinese, has a three-way contrast of unaspirated voiceless, aspirated voiceless, and

voiced stops. Pan found that the two girls acquired unaspirated voiceless stops first, voiceless aspirated stops second, and voiced stops last. In addition, the two girls could produce the three-way contrast by three years of age, but not in an adult-like manner.

Given the scarcity of research on the acoustic development of stop consonants in Mandarin-speaking children, the present study will fill the knowledge gap by documenting and investigating the development of temporal features of stop consonants in three- to six-year-old Mandarin-speaking children relative to Mandarin-speaking adults. Because Mandarin stops are characterized by a short-lag and long-lag VOT distinction, similar to English stops, we expect that Mandarin-speaking children will show a delayed acquisition of long-lag VOTs in comparison to short-lag VOTs, similar to English-speaking children. That is, they will show more difficulty producing the aspirated stops in an adult-like manner. Meanwhile, by virtue of the universal trend of developing coordination of respiration, phonation, and articulation from children to adults, regardless of the children's language backgrounds, we expect that Mandarin children will also experience continuing development of the durational features associated with the physiological constraints. Therefore, they will show different patterns from the adults on the durational measurements of stop components.

## Methods

### Participants

The participants included 29 native Mandarin-speaking children (13 females and 16 males) aged three to six years old, and 12 native Mandarin-speaking adults (6 females and 6 males) aged 23 to 58 years old. All children were born in Mandarin-speaking regions and raised in the Beijing area by Mandarin-speaking parents. The 29 children included 6 three-year-olds (3 females, 3 males,  $M = 3;6$ ,  $SD = 5.4$  mo), 8 four-year-olds (3 females, 5 males,  $M = 4;6$ ,  $SD = 3.0$  mo), 11 five-year-olds (4 females, 7 males,  $M = 5;8$ ,  $SD = 3.7$  mo), and 4 six-year-olds (3 females, 1 male,  $M = 6;4$ ,  $SD = 2.2$  mo). The adult speakers ( $M = 34$  yr,  $SD = 12$  yr) were born and raised in Mandarin-speaking regions. Five of them were teachers employed in the kindergarten where the children were recruited. No speech, language, or hearing impairments were reported by any of the participants, their parents, or kindergarten teachers.

### Speech materials

The speech materials included 18 Mandarin disyllabic words containing six Mandarin stops /p, t, k, p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/, each followed by three vowels /a, i, u/ (see Table 2 for the word list). The three vowels /a, i, u/ occupy the corner positions of the Mandarin vowel space and are commonly used as the neighboring vowel contexts for stop studies in Mandarin adults (Chao & Chen, 2008; Chen *et al.*, 2007). Due to the phonotactic constraints of Mandarin, velar stops do not occur in the syllables with the vowels /i/ or /y/. The vowel /i/ was substituted with /ɿ/. For all 18 words, the target stop consonants always occurred in word-initial position in the first syllable. The selection of the words followed the general rule that all words should be easily presented in pictures which were recognizable and familiar to children in this age range. The tone

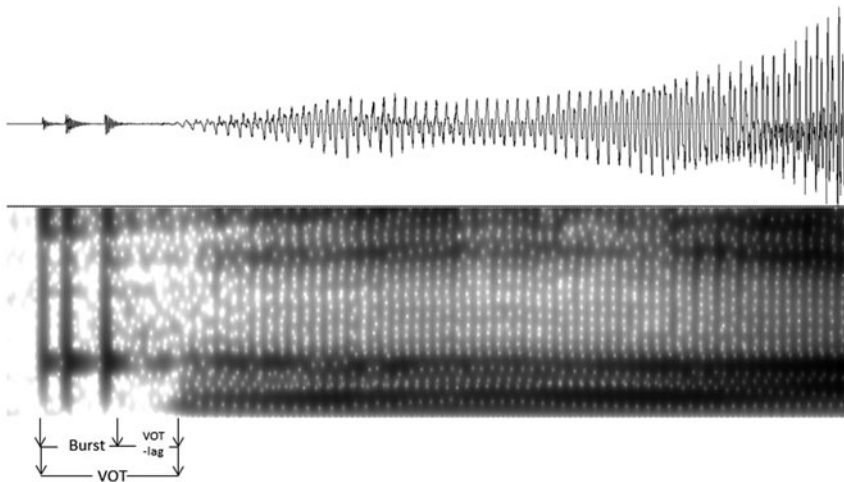
**Table 2.** The Word List Used for Data Collection.

Stop	Pinyin	IPA	Gloss
p	bà ba	/papa/	father
	bí zi	/pitsɿ/	nose
	bō luó	/puoluo/	pineapple
p <sup>h</sup>	pán zi	/p <sup>h</sup> antsɿ/	plate
	pí qiú	/p <sup>h</sup> itɕ <sup>h</sup> ioʊ/	ball
	pú táo	/p <sup>h</sup> ut <sup>h</sup> au/	grapes
t	dà suàn	/tasuan/	garlic
	dì di	/titi/	younger brother
	dù zi	/tutsɿ/	belly
t <sup>h</sup>	táng guǒ	/t <sup>h</sup> aŋkuo/	candy
	tī zi	/t <sup>h</sup> itsɿ/	ladder
	tǔ dòu	/t <sup>h</sup> utoʊ/	potato
k	gài zi	/kartsɿ/	lid
	gē zi	/kɤtsɿ/	pigeon
	gú tou	/kut <sup>h</sup> oʊ/	bone
k <sup>h</sup>	kǎ chē	/k <sup>h</sup> atɕ <sup>h</sup> ɤ/	truck
	kě lè	/k <sup>h</sup> ɤtsɿ/	Coca Cola
	kù zi	/k <sup>h</sup> utsɿ/	pants

environment was not strictly controlled because the picturability and familiarity of the target words restricted word selection. Moreover, the effect of lexical tone on the word-initial voiceless stops was limited because the consonant–tone interaction was primarily manifested between voiced consonants and low tones (Bradshaw, 1999), which was not of concern in the present study.

### Recording

The recordings were conducted in a quiet room or a sound-attenuated room for children and adult participants by a trained experimenter. Each participant was seated in front of a laptop computer and was recorded producing the target words through a visual–auditory word repetition task under the control of a custom MATLAB program. A word repetition task was used to ensure a consistent phonetic environment for the target words across all speakers and to ensure that the speech samples were elicited in the designed manner (Edwards & Beckman, 2008). The recording session included two blocks. In each block, pictures representing the 18 target words were randomly ordered and shown on the computer screen. For each target word, the participants first heard an audio prompt produced in a citation form by a native Mandarin speaker and were then asked to repeat the word once immediately after the audio prime. Speech samples were recorded through a Shure



**Figure 1.** The waveform and spectrogram of the token *ge* produced by one child speaker. The word-initial stop /k/ was produced with multiple bursts.

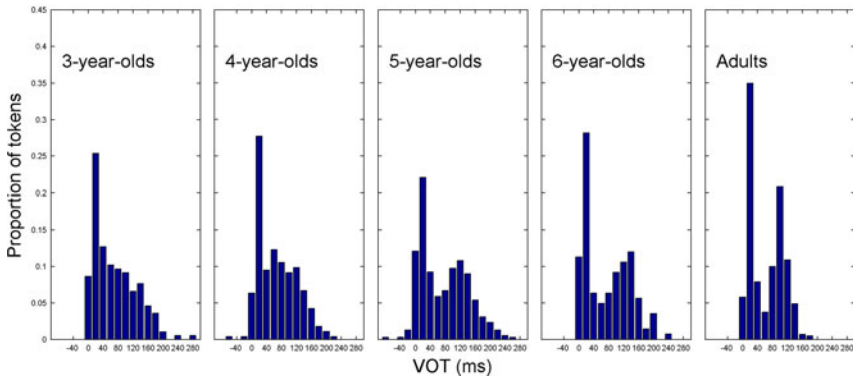
SM10A head-mounted microphone placed approximately one inch from the speaker's mouth. All data were recorded directly onto the hard disk of a laptop with a 16-bit quantization rate and 44.1 kHz sampling rate. Each participant produced 36 tokens (18 words  $\times$  2 blocks), which resulted in a total of 1,476 tokens. However, due to skipped tokens during the recording process or the absence of syllable-initial stop sounds, the total number of 1,446 tokens were used for further analysis.

### *Data analysis*

The landmarks of the onset and offset of the release burst and the onset of vocal fold vibration were determined by an experienced researcher through observing the spectrogram with the assistance of waveform using the spectrographic analysis program Adobe Audition 1.0. For the stops with multiple bursts (shown in Figure 1), the onset and offset were marked for each burst. VOT was measured as the time interval from the onset of the first release burst to the onset of voicing. For the majority of stops that were produced as true voiceless sounds, the onset of voicing was represented as the onset of the vowel following the target stop sound. However, there were also a few tokens that were produced with a voiced stop at the word-initial position. For those exceptions, the onset of voicing was measured separately from the onset of the following vowel.

In addition to the VOT, we also examined the overall burst duration, average duration per burst, number of bursts, and VOT-lag duration. The overall burst duration was defined as the duration from the beginning of the first burst to the end of the last burst. To calculate the average duration per burst, the sum of burst duration for all bursts in a given token was first obtained, and was then divided by the number of bursts. For those stops with only a single burst, the overall burst duration was equal to the average duration per burst. The VOT-lag was defined as the interval between the end of the last burst and the onset of the following vowel.





**Figure 2.** Histogram showing the distribution of VOT data for all six Mandarin stops in three-, four-, five-, and six-year-old children and the adults.

Some unaspirated stops have the onset of the following vowel occurring immediately after the release burst. In this case, there is no delay of vocal fold vibration and the VOT-lag is 0. Note that if the unaspirated stop has a single burst, the VOT, overall burst duration, and average duration per burst have the same value. In other cases, the VOT is the sum of the overall burst duration and VOT-lag duration. In aspirated stops, the duration of VOT-lag is the duration of aspiration. A reliability check was implemented by another trained experimenter who checked the landmark locations of the onsets and offsets of stop components for all tokens. A 96.6% consistency rate was found between the two experimenters. Any disagreement and random errors were resolved between the two experimenters following discussion.

To test whether the durational measurements differ between males and females, a three-way repeated measures ANOVA was conducted on these measurements of the 12 adults with gender as the between-subject factor and stop and vowel as the within-subject factors. The results revealed no significant gender effect. Regarding the children's speech, previous studies suggested that the articulatory structure does not show apparent differences between genders until around seven years of age (Vorperian, Kent, Lindstrom, Kalina, Gentry, & Yandell, 2005). The children in the present study were all younger than seven years old. The effect of gender difference on speech acoustic features for these children was expected to be limited. Repeated measures ANOVAs were conducted on the temporal measurements of the 29 children with gender and age as the between-subject factors, and stop and vowel as the within-subject factors. The results revealed no significant gender differences. Based on these preliminary analyses, the data were collapsed across gender for further examination.

## Results

### *VOT development in Mandarin-speaking children*

Figure 2 shows the distribution of all VOTs plotted for three-, four-, five-, and six-year-old children and adult speakers. The purpose of this analysis was to provide an overall profile of VOTs in each age range so that a general developmental trend could be observed. The adults demonstrated a well-shaped bimodal distribution with

the VOT values gathering at a short-lag region below 40 ms and a long-lag region around 100 ms. All children showed high occurrences of short-lag VOTs between 0 and 20 ms, similar to the adult speakers. However, the distribution of long-lag VOTs in the children was not as concentrated as that in the adults. Specifically, the three- and four-year-olds had a similar widespread distribution of VOTs between 40 and 160 ms. In the five- and six-year olds, the VOT values showed a gradually increased concentration between 100 and 140 ms, in addition to the concentration at around 20 ms. The VOT distribution demonstrated a unimodal pattern in the three- and four-year-olds, which developed into an emerging bimodal pattern in the five- and six-year-olds and gradually approximated the bimodal distribution of the adult speakers.

Considering that the number of child participants was not equally allocated in each age group, and that the overall distribution of the durational measurements (including VOTs and other measures) demonstrated a relatively consistent pattern between the three- and four-year-olds, as well as between the five- and six-year-olds, the children were reorganized into two groups for further analysis. Specifically, the three- and four-year-olds were grouped into younger children (AY), and the five- and six-year-olds were grouped into older children (AO) to be compared with the adults (AA). To better examine whether the Mandarin-speaking children at this age have acquired an adult-like timing pattern for the unaspirated and aspirated stops, the durational measurements were separately plotted and analyzed for unaspirated and aspirated stops in the younger children, older children, and adults. Three-way repeated measures ANOVAs were conducted on each durational measure, with the place of articulation and vowel context as the within-subject factors and age as the between-subject factor. A Bonferroni test with adjustment for multiple comparisons was used for subsequent pairwise comparisons when a significant main effect was yielded. Simple contrasts were also used to compare different levels of the within-subject factors and for the post-hoc analysis for the interaction effects.

Figure 3 shows the distribution of all VOT data (top), and VOT data for the unaspirated stops (middle) and aspirated stops (bottom) in the younger children, older children, and adults. The overall VOT distribution showed a gradually developing separation between the short-lag and long-lag VOTs from the younger children to the adults. As for the unaspirated stops, over 70% of VOTs for Mandarin /p, t, k/ were located between 20 and 40 ms in the adults. Both groups of children produced the majority of unaspirated stops with short-lag VOTs similar to the adults. However, there were sporadic cases in which some children produced certain intended unaspirated stops with long-lag VOTs over 100 ms. For the aspirated stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/, the adults showed a high concentration of VOTs around the 100 ms region. Both groups of children exhibited a wider range and a much less concentrated distribution of the VOT values than the adults.

To better present the VOT variations related to age difference from children to adults, Figure 4 plots the means and standard errors of the VOT data collapsed across the places of articulation and vowel contexts in each group of speakers. For all six stop consonants, especially the aspirated stops, the two groups of children produced longer VOTs than the adults. In particular, the difference between the older children and the adults was larger than the difference between the younger children and the adults for the aspirated stops. Table 3 lists the means and standard errors of VOT values for each stop in each vowel context in the younger children, older children, and adults. Similar to the adults, both groups of children produced

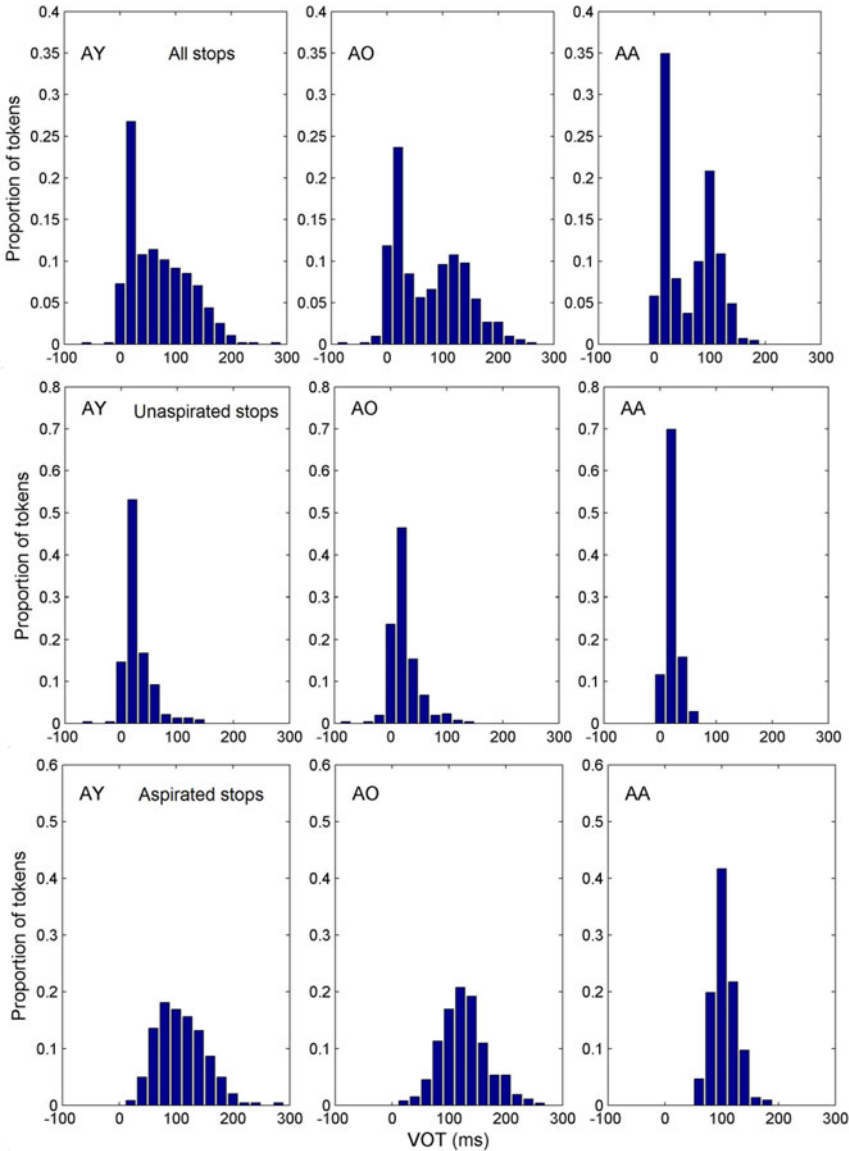
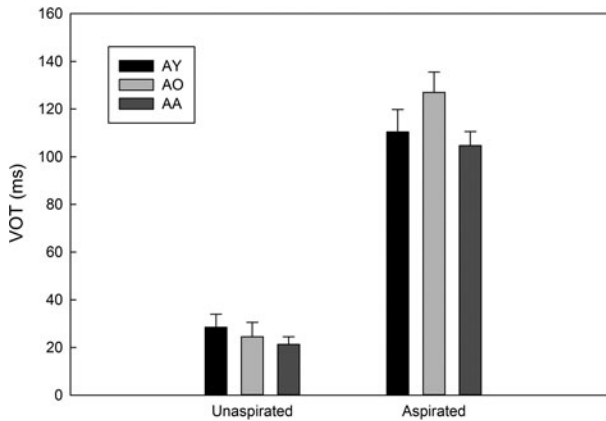


Figure 3. Histogram showing the distribution of VOT data for all stops (top), unaspirated stops (middle) and aspirated stops (bottom) in the younger children, older children, and adults.

velar stops with longer VOT than bilabial or alveolar stops. In addition, they tended to produce the stops in the vowel context /a/ with shorter VOTs than in the other vowel contexts. The three-way repeated measures ANOVA tests yielded a significant age effect on the aspirated stops ( $F(2,38) = 5.168, p = .01$ ) but not on the unaspirated stops. A subsequent pairwise comparison revealed a significantly longer VOT in the older children than in the adults for the aspirated stops. Regarding the main effects of



**Figure 4.** Bar plot showing the overall means and standard errors of VOTs for unaspirated and aspirated stops collapsed across the three places and vowel contexts in the younger children, older children, and adults.

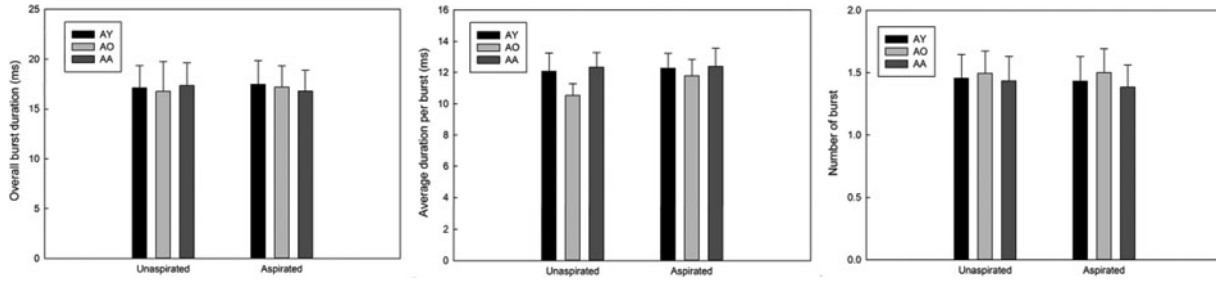
place and vowel context, the ANOVA test revealed a significant main effect of place for both unaspirated ( $F(2,76) = 39.650, p < .0001$ ) and aspirated stops ( $F(2,76) = 3.853, p = .025$ ). The pairwise comparisons showed that VOT values were longer in the velar stops than in the other two places. The vowel context was found to be a significant main effect for both unaspirated ( $F(2,76) = 39.801, p < .0001$ ) and aspirated stops ( $F(2,76) = 35.184, p < .0001$ ). Specifically, the VOTs of stops followed by the vowel /a/ were significantly shorter than those followed by the other vowels. No interaction effect of age by place or age by vowel was found for the unaspirated or aspirated stops, which suggested that the children showed varying VOT values as a function of place and vowel context in a pattern similar to the adults. A place by vowel interaction effect was significant for both unaspirated ( $F(4,152) = 9.080, p < .0001$ ) and aspirated stops ( $F(4,152) = 6.247, p < .0001$ ). This interaction effect suggested that the effect of the subsequent vowel on VOT values was not consistent across different places of articulation.

#### *Durational measurements of stop components*

In addition to VOT, the present study also compared the temporal features associated with stop components of the burst (shown in Figure 5 and Table 4) and VOT-lag (shown in Figure 6 and Table 5). The left panel of Figure 5 shows the age difference of overall burst duration collapsed across articulation places and vowel contexts. While the two groups of children tended to produce longer overall burst duration than the adults for the aspirated stops, there was no observable difference between the younger children and the adults for the unaspirated stops. The means and standard errors of the overall burst duration of each stop in each vowel context for all three groups are listed in Table 4. It is clearly shown that both children and adults produced longer burst duration for velar stops than for the other two places. The results of the repeated measures ANOVA revealed that there was no significant age difference for the unaspirated or aspirated stops, but that there was a significant main effect of place in both unaspirated ( $F(2,76) = 107.583, p < .0001$ ) and aspirated

**Table 3.** Means and Standard Errors (in ms) of VOT for Each Stop in Each Vowel Context in the Younger Children (AY), Older Children (AO), and Adults (AA).

		pa	pi	pu	p <sup>h</sup> a	p <sup>h</sup> i	p <sup>h</sup> u	ta	ti	tu
AY	M	11.89	21.93	37.57	99.00	114.46	117.39	16.89	35.79	25.00
	SE	1.68	4.88	6.33	8.46	9.25	9.79	3.40	6.63	5.63
AO	M	9.27	15.97	33.43	110.27	142.00	133.33	14.53	13.37	13.37
	SE	2.09	1.67	5.73	8.13	8.87	9.17	2.59	6.60	2.33
AA	M	11.75	14.42	24.58	98.50	102.21	105.46	13.04	17.88	15.25
	SE	0.72	1.46	2.53	4.11	5.37	4.49	0.98	1.12	0.93
		t <sup>h</sup> a	t <sup>h</sup> i	t <sup>h</sup> u	ka	kɿ	ku	k <sup>h</sup> a	k <sup>h</sup> ɿ	k <sup>h</sup> u
AY	M	99.71	127.29	100.82	21.64	36.07	49.89	90.25	118.43	126.32
	SE	10.04	9.31	9.75	1.93	3.31	5.36	5.35	6.77	11.84
AO	M	103.20	137.67	118.33	25.53	44.07	51.00	109.03	138.13	151.37
	SE	4.98	8.09	5.74	6.05	6.16	6.80	6.60	6.59	9.10
AA	M	94.42	112.46	96.33	21.71	33.67	39.75	96.00	116.83	120.13
	SE	4.30	8.70	5.17	1.68	2.51	3.13	3.82	6.07	5.35



**Figure 5.** Bar plots showing the overall means and standard errors of burst-related measurements (overall burst duration, average duration per burst, and number of bursts) for unaspirated and aspirated stops collapsed across the three places and vowel contexts in the younger children, older children, and adults.

**Table 4.** Means and Standard Errors (in ms) of the Burst Features Including the Overall Burst Duration, Average Duration per Burst, and Number of Bursts for Each Stop in Each Vowel Context in the Younger Children (AY), Older Children (AO), and Adults (AA).

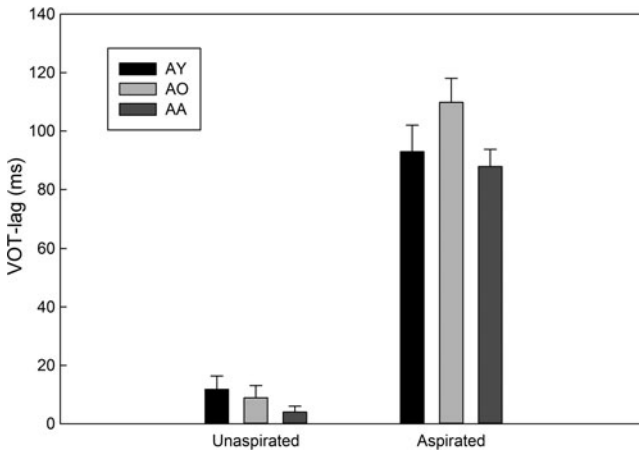
			pa	pi	pu	p <sup>h</sup> a	p <sup>h</sup> i	p <sup>h</sup> u	ta	ti	tu
Overall burst duration	AY	M	10.64	12.25	16.79	11.71	11.75	16.68	12.36	15.04	14.71
		SE	0.76	1.16	1.70	0.62	0.68	2.10	1.18	1.17	0.95
	AO	M	8.87	11.83	14.67	12.33	13.60	13.13	12.03	13.17	12.50
		SE	0.38	1.33	1.13	0.67	1.01	1.05	0.95	1.07	0.69
	AA	M	11.75	10.63	15.17	13.83	11.00	15.00	12.79	15.63	14.13
		SE	0.72	0.65	1.45	0.87	0.54	1.66	0.84	0.79	0.59
Average duration per burst	AY	M	10.07	10.55	11.60	11.05	11.45	11.58	9.98	14.36	13.88
		SE	0.72	0.51	1.14	0.50	0.71	0.96	0.71	1.17	1.08
	AO	M	8.43	10.28	9.45	10.63	12.50	10.26	9.61	11.45	11.42
		SE	0.32	0.59	0.66	0.66	1.12	0.65	0.74	1.07	0.71
	AA	M	11.42	9.65	11.38	11.97	10.46	11.59	11.40	15.40	14.13
		SE	0.68	0.51	1.02	0.49	0.49	1.25	0.81	0.84	0.59
Number of bursts	AY	M	1.07	1.11	1.50	1.11	1.04	1.32	1.32	1.04	1.11
		SE	0.07	0.06	0.10	0.08	0.04	0.15	0.22	0.04	0.06
	AO	M	1.07	1.10	1.37	1.23	1.13	1.23	1.33	1.27	1.13
		SE	0.05	0.07	0.10	0.08	0.08	0.11	0.11	0.10	0.06
	AA	M	1.04	1.13	1.38	1.21	1.04	1.29	1.17	1.04	1.00
		SE	0.04	0.07	0.16	0.11	0.04	0.14	0.09	0.04	0.00
			t <sup>h</sup> a	t <sup>h</sup> i	t <sup>h</sup> u	ka	k <sub>s</sub>	ku	k <sup>h</sup> a	k <sup>h</sup> <sub>s</sub>	k <sup>h</sup> u
Overall burst duration	AY	M	14.36	13.54	14.11	16.68	25.25	30.36	18.96	32.36	23.89
		SE	0.75	0.84	0.75	0.73	1.93	2.90	1.59	2.68	2.87

(Continued)

Table 4. (Continued.)

			t <sup>h</sup> a	t <sup>h</sup> i	t <sup>h</sup> u	ka	k <sub>v</sub>	ku	k <sup>h</sup> a	k <sup>h</sup> v	k <sup>h</sup> u
	AO	M	13.77	14.00	12.70	22.40	26.07	29.30	20.97	27.27	27.03
		SE	0.74	1.43	0.69	5.50	2.35	3.66	2.13	2.33	2.39
	AA	M	16.00	10.63	12.38	18.67	29.33	28.08	19.17	27.25	25.92
		SE	1.01	0.47	0.44	1.62	2.26	1.92	1.51	2.04	1.81
Average duration per burst	AY	M	12.69	12.27	14.11	12.59	11.97	13.67	11.63	11.87	13.87
		SE	0.73	0.92	0.75	0.77	0.77	2.32	0.64	1.18	1.56
	AO	M	12.83	11.28	12.70	12.37	10.63	11.18	14.43	9.83	11.72
		SE	0.54	0.77	0.69	0.94	0.56	0.59	1.82	0.76	1.39
	AA	M	13.27	10.63	12.38	13.17	10.71	13.91	13.09	16.01	12.08
		SE	0.62	0.47	0.44	1.16	0.63	0.83	1.59	2.29	1.03
Number of bursts	AY	M	1.18	1.18	1.00	1.43	2.18	2.36	1.64	2.71	1.71
		SE	0.08	0.14	0.00	0.10	0.19	0.29	0.14	0.27	0.18
	AO	M	1.07	1.30	1.00	1.87	2.20	2.13	1.57	2.80	2.17
		SE	0.05	0.12	0.00	0.29	0.14	0.20	0.17	0.21	0.20
	AA	M	1.17	1.00	1.00	1.54	2.54	2.08	1.71	1.92	2.13
		SE	0.07	0.00	0.00	0.16	0.23	0.20	0.16	0.25	0.21





**Figure 6.** Bar plot showing the overall means and standard errors of VOT-lag duration for unspirated and aspirated stops collapsed across the three places and vowel contexts in the younger children, older children, and adults.

stops ( $F(2,76) = 119.129, p < .0001$ ). The pairwise comparisons revealed that the overall burst durations of velar stops were longer than those of bilabial and alveolar stops. In addition to the main effect of place, there was a significant main effect of vowel context on the overall burst duration for both unspirated ( $F(2,76) = 17.322, p < .0001$ ) and aspirated stops ( $F(2,76) = 8.833, p < .0001$ ). The pairwise comparison revealed that the bursts of the stops followed by /a/ were shorter than those of the stops followed by the other vowels. No interaction effects of age by place or age by vowel were found for the unspirated or aspirated stops, which suggested that the children showed a pattern of overall burst duration as a function of place and vowel context similar to the adults. A place by vowel interaction effect was significant for both unspirated ( $F(4,152) = 5.134, p = .006$ ) and aspirated stops ( $F(4,152) = 14.816, p < .0001$ ). These results indicated that the effect of vowel context on the overall burst duration was inconsistent across different places of articulation.

The middle panel of [Figure 5](#) presents the age difference of the average duration per burst collapsed across vowel contexts and places of articulation. The adults produced longer average duration per burst than the older children but did not show much difference from the younger children for both unspirated and aspirated stops. The means and standard errors of the average duration per burst for each stop in each vowel context for all three groups of speakers are presented in [Table 4](#). The ANOVA results showed a significant age effect for the unspirated stops ( $F(2,38) = 9.204, p = .001$ ) but not for the aspirated stops. In particular, the older children produced significantly shorter average duration per burst for the unspirated stops than the younger children and the adults. In addition, the main effect of place was found to be significant for both unspirated ( $F(2,76) = 11.316, p < .0001$ ) and aspirated stops ( $F(2,76) = 3.631, p = .040$ ). The pairwise comparison post-hoc analysis revealed significantly shorter average duration per burst for the bilabial stops than the alveolar and velar stops. The main effect of vowel was found to be significant in the unspirated stops ( $F(2,76) = 6.415, p = .003$ ) but not in the aspirated stops. The average duration per burst of the unspirated stops followed by the vowel /a/ was

**Table 5.** Means and Standard Errors (in ms) of VOT-lag Duration for Each Stop in Each Vowel Context in the Younger Children (AY), Older Children (AO), and Adults (AA).

		pa	pi	pu	p <sup>h</sup> a	p <sup>h</sup> i	p <sup>h</sup> u	ta	ti	tu
AY	M	2.00	12.29	20.79	87.29	102.71	100.71	4.54	20.75	10.29
	SE	1.25	3.65	5.80	8.60	8.95	9.02	3.00	6.62	5.55
AO	M	1.63	4.13	19.50	97.93	128.40	120.20	2.50	6.83	2.83
	SE	1.63	1.39	5.83	8.25	8.57	8.87	1.89	3.37	1.22
AA	M	0.00	3.79	9.42	84.67	91.21	90.46	0.25	2.25	1.13
	SE	0.00	1.50	2.83	4.35	5.49	4.95	0.25	1.03	1.12
		t <sup>h</sup> a	t <sup>h</sup> i	t <sup>h</sup> u	ka	k <sub>ɻ</sub>	ku	k <sup>h</sup> a	k <sup>h</sup> ɻ	k <sup>h</sup> u
AY	M	85.32	113.75	86.71	4.93	10.82	19.54	71.29	86.07	102.43
	SE	9.70	9.14	9.78	1.72	3.19	5.22	5.25	6.16	10.96
AO	M	89.43	123.67	105.63	3.13	18.00	21.63	88.07	110.87	124.27
	SE	5.01	7.12	5.79	2.49	5.82	5.88	7.22	7.62	8.70
AA	M	78.42	101.83	83.96	3.04	4.33	11.67	76.83	89.58	94.21
	SE	4.73	8.49	5.04	1.29	1.95	3.04	4.53	6.69	5.09

significantly shorter than that followed by the vowel /u/. Regarding the interaction effects, there was a significant place by vowel effect ( $F(4,152) = 5.964, p < .0001$ ), but no significant age-related two-way or three-way interaction effects for the unaspirated stops. For the aspirated stops, there was a significant age by vowel ( $F(4,76) = 3.110, p = .020$ ) interaction effect. In particular, the three groups of speakers showed different patterns of average duration per burst for the aspirated stops followed by /a/ from those followed by /u/. A significant place by vowel by age interaction effect ( $F(8,152) = 2.971, p = .004$ ) was also found for aspirated stops, which suggested that the three groups of speakers showed different average duration per burst in certain places and vowel contexts.

The right panel of Figure 5 displays the age difference of the number of bursts collapsed across articulation places and vowel contexts. Both groups of children, especially the older children, showed a tendency of a greater number of bursts than the adults. This means that multiple bursts tended to occur more frequently in the children than in the adults. However, as this age-related difference was subtle, no significant age effect was found for the unaspirated or aspirated stops. According to the means of number of bursts for each stop in different vowel contexts shown in Table 4, both children and adults produced a greater number of bursts for the velar stops /k, k<sup>h</sup>/ than for the bilabial and alveolar stops. This pattern reflected an articulatory constraint in producing stop consonants in various places (Maddieson, 1997; Stevens, 2000). The statistical analysis confirmed the significant effect of place of articulation on the number of bursts for both unaspirated ( $F(2,76) = 87.831, p < .0001$ ) and aspirated stops ( $F(2,76) = 77.844, p < .0001$ ). In addition, the main effect of vowel was also significant for both unaspirated ( $F(2,76) = 6.859, p = .003$ ) and aspirated stops ( $F(2,76) = 9.641, p < .0001$ ). The number of bursts in stops followed by /a/ was smaller than that in stops followed by the other vowels. Regarding the interaction effects, there was a significant place by vowel effect in both unaspirated ( $F(4,152) = 10.831, p < .0001$ ) and aspirated stops ( $F(4,152) = 17.767, p < .0001$ ). This interaction effect suggested that the number of bursts for different places was affected by the following vowel context. In addition, there was a significant vowel by age interaction effect ( $F(4,76) = 4.475, p = .003$ ) for the aspirated stops. In particular, the three groups showed different patterns of the number of bursts for the stops followed by /i/ than for those followed by /a/ or /u/. Moreover, the aspirated stops also demonstrated a significant place by vowel by age interaction effect ( $F(8,152) = 2.832, p = .007$ ), which suggested that the children showed different patterns of the number of bursts for the three places of articulation in different vowel contexts.

Figure 6 displays the age difference of VOT-lag duration collapsed across vowel contexts and places of articulation. VOT-lag is defined as the interval between the end of the burst and the onset of the following vowel. Note that VOT-lag represents the aspiration following the release burst in aspirated stops. For both unaspirated and aspirated stops, the two groups of children produced longer VOT-lags than the adults. Table 5 shows the mean durations of VOT-lags for each stop in each vowel context for all three groups of speakers. For the unaspirated stops, the ANOVA results yielded a significant main effect of age ( $F(2,38) = 5.073, p = .011$ ), which was mainly represented as a longer VOT-lag in the younger children than in the adults. For the aspirated stops, the ANOVA results also yielded a significant age effect ( $F(2,38) = 5.399, p = .009$ ), which was mainly represented as a significantly longer VOT-lag in the older children than in the younger children and adults. The main

effect of place was significant in the unaspirated stops ( $F(2,76) = 3.838, p = .032$ ) but not in the aspirated stops. The pairwise comparison showed that the unaspirated velar stop /k/ had a longer VOT-lag than the unaspirated alveolar stop /t/. The main effect of vowel was found to be significant in both unaspirated ( $F(2,76) = 24.001, p < .0001$ ) and aspirated stops ( $F(2,76) = 29.835, p < .0001$ ). The pairwise comparisons showed that the duration of VOT-lag in the stops followed by /a/ was shorter than that in stops followed by the other vowels. In terms of the interaction effects, no significant age-related two-way or three-way interaction effects were found for the unaspirated or aspirated stops. These results indicated that the children followed the adults in the pattern of VOT-lag duration across places of articulation and vowel contexts. However, the place by vowel interaction effect was significant for both unaspirated ( $F(4,152) = 5.304, p = .001$ ) and aspirated stops ( $F(4,152) = 5.754, p < .0001$ ). This interaction effect indicated that the effect of articulation place on VOT-lag duration was inconsistent across different vowel contexts.

## Discussion

Many previous studies have investigated the development of stop consonants in English-speaking children younger than three years of age (e.g., Imbrie, 2005; Lowenstein & Nittrouer, 2008; Macken & Barton, 1980a). The present study, examining VOT and other temporal features of stop components in Mandarin-speaking children aged three to six years old, has provided additional evidence to show the continuing development of stop consonants in children in this age range.

Previous studies on the VOT development of English-speaking children have reported that short-lag VOTs emerged and were acquired earlier than long-lag VOTs. For example, Gilbert (1977) found that English-speaking children aged 2;7–3;3 had acquired an adult-like short-lag category, but showed a widely dispersed distribution of long-lag VOTs. In a more recent study, Lowenstein and Nittrouer (2008) reported that English-speaking children aged between 1;2 and 2;7 produced voiced stops with adult-like short-lag VOTs, but that the long-lag VOTs for the voiceless stops were not as long as the adult targets. In the present study, we observed that Mandarin-speaking children at three years old had well-developed short-lag VOTs, but that the Mandarin-speaking children as old as six years of age had not fully established an adult-like long-lag VOT model. This observation indicates the continuing development of long-lag VOTs in Mandarin-speaking children in this age range. Compared to highly clustered adult-like short-lag VOTs, long-lag VOTs were acquired later in Mandarin-speaking children, similar to English-speaking children.

Zlatin and Koenigsnecht (1976) generalized that the VOT distribution in English-speaking children changed from a unimodal to an emerging bimodal to a typical bimodal pattern. In the present study, the three- to six-year-old Mandarin-speaking children also showed a similar developmental path. However, considering the age difference between the participants in the current study and those in previous developmental studies, we should not directly correlate the development stages in the present study to the developmental paths proposed in previous studies. Previous studies reported that English-speaking children experienced a rapid development of VOT and demonstrated increased approximation to the adults' values at around three years of age (Lowenstein & Nittrouer, 2008; Macken &

Barton, 1980a). Therefore, most phonetic studies on stop acquisition in English-speaking children focused on children at a young age (normally no older than three years old). They found that English-speaking children at a very young age (younger than two years old) predominantly produced short-lag VOTs and avoided long-lag VOTs in the unimodal stage (Lowenstein & Nittrouer, 2008; Macken & Barton, 1980a; Zlatin & Koenigsknecht, 1976). The target population in the present study were aged between three and six years old. The younger children (three to four years of age) indeed produced long-lag VOTs, which was represented as a similar overall VOT range in children and adults. However, they did not show as well-concentrated a distribution of long-lag VOTs as that shown in the adult speakers. Given that Mandarin aspirated stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/ are normally produced with a strong air flow and are characterized by a longer VOT than the English voiceless stops /p, t, k/, the continuing development of long-lag VOTs in relatively older Mandarin-speaking children provides additional evidence showing that the production of long-lag VOTs requires more delicate timing control and articulatory coordination than the production of short-lag VOTs (Gilbert, 1977; Imbrie, 2005; Kewley-Port & Preston, 1974).

To better examine the development of the temporal pattern of articulatory coordination from children to adults, the present study further compared the overall burst duration, average duration per burst, number of bursts, and VOT-lag duration between children and adults. Although no age difference was found for the overall burst duration among the three groups of speakers, the children produced shorter average duration per burst than the adults. The statistical analysis revealed significantly shorter average duration per burst in the older children than in the adults for the unaspirated stops, and age-related interaction effects for the aspirated stops. Imbrie (2005) stated that the shorter average duration per burst in children relative to adults reflected the immature respiration control and higher subglottal pressure in children. The present study provided further evidence in favor of this notion. Imbrie found a general decreasing trend in the number of bursts from children to adults for all three places of articulation, and the age difference was statistically significant. The present study did not show a significant age difference in the average number of bursts. However, both groups of children, especially the older children, showed an observable greater number of bursts than the adults. This observation evidenced the higher incidence of multiple bursts in children relative to adults, which can be explained by the increased compliance of the articulator or high subglottal pressure during the occlusion release in children (Imbrie, 2005).

In addition to the burst related measurements, these children produced significantly longer VOT-lags than the adults for both unaspirated and aspirated stops. As a measurement indexing the coordination of oral and laryngeal articulatory gestures (Imbrie, 2005; Löfqvist, 1980, 1992), the significantly longer VOT-lag in these children suggests that children in this age may not have established adult-like patterns of oral-laryngeal coordination. It is noteworthy that these children did not show a significant difference from the adults on the short-lag VOTs for the unaspirated stops. However, they produced shorter average duration per burst and longer VOT-lag for the unaspirated stops than the adults. This finding indicates that even though these children might have developed an adult-like distribution for the short-lag VOTs for the unaspirated stops, they might still differ from the adults in the mechanism of articulatory coordination for the unaspirated stops. The immature inter-articulator timing in children was also partially evidenced by the negative VOT

values observed in certain speakers from both groups of children. Unlike English voiced stops that are manifested as negative VOT values in certain regional dialects or speakers, all six Mandarin stop consonants are voiceless stops produced with positive VOT values. In the present study, no single negative VOT value was present in the adult speakers. However, some children showed a few cases of negative VOT for the voiceless stops (as shown in [Figure 2](#)). These results indicated that some of the children initiated the vocal fold adduction prior to the release of oral closure. Therefore, children in this age range may not have established the adult-like sequential relationship between the laryngeal and oral gestures.

While previous developmental studies agreed that children normally differed from adults in one or more aspects of stop features, it remains controversial whether children produced longer or shorter VOTs than the adults. Some researchers reported that children produced shorter VOT means than adults (Kewley-Port & Preston, 1974; Zlatin & Koenigsnecht, 1976). Other researchers found evidence for longer VOT means in young children than in adults (Imbrie, 2005; Barton & Macken, 1980; Smith, 1978). In the present study, our data showed that Mandarin-speaking children aged three to six years old produced all six stops, especially the aspirated stops, with longer VOTs than the adults. When comparing the durational and burst measurements in the two groups of children with the adults, the results showed that the difference between the older children and the adults was even larger than the difference between the younger children and the adults. This finding suggested that the VOT values might ‘overshoot’ from the younger children to the older children and then gradually shorten back to adult values as the children’s timing control develops. The overshooting of VOT values in children relative to adults was also observed in earlier studies involving children from different language backgrounds even though the age of participants in different studies varied (Barton & Macken, 1980; Lee & Iverson, 2008, 2012, 2017). Lee and Iverson (2012, 2017) examined the development of phonetic categories in bilingual Korean–English children and corresponding monolingual children. In these studies, the authors reported the VOT data in three-, five-, and ten-year-old English monolingual children. Similar to the overshooting VOT values observed in the present study, the VOTs for the voiceless stops in the five-year-olds were longer than those in the three-year-olds and shortened in the ten-year-olds. These findings together demonstrate that the acoustic development of stop consonants is a protracted process which extends beyond three years of age.

In addition to the developmental path of stop consonants from young children to adults, the VOT values of the adults in the present study were similar to, and showed a compatible pattern with, the VOT means reported for Mandarin Chinese (e.g., Rochet & Fei, 1991). Previous studies also showed that, regardless of the language being tested, VOT tends to be longer as the place of constriction moves from front to back and as the following vowel context changes from low to high (Chen *et al.*, 2007; Cho & Ladefoged, 1999; Morris, McCrea, & Herring, 2008; Rochet & Fei, 1991; Weismer, 1979). The present study provides additional evidence supporting these findings. Moreover, in the Mandarin speakers tested, velar stops were produced with longer overall burst duration, average duration per burst, and VOT-lag duration as well as a greater number of bursts than other places. Stops followed by the low vowel /a/ showed shorter durational features and a smaller number of bursts than the stops followed by high vowels. These findings suggest that the detailed acoustic features of stop components, including stop burst and VOT-lag,

are subject to the impact of vowel context and place of articulation, in a similar manner to VOT, in Mandarin as in English. For some of the measures tested, the children displayed varying temporal measurements as a function of place and vowel context similar to those of the adults. This finding revealed the similar articulatory placement of stop production in children relative to adults.

In sum, the present study documented the continuing acoustic development of stop consonants in three- to six-year-old Mandarin-speaking children. Children in this age range produced adult-like short-lag VOTs, but their long-lag VOTs were longer than the adults. Even though both groups of children produced adult-like short-lag VOTs, this did not mean that they produced the unaspirated stops in an adult-like manner. Further analysis of the stop components revealed that the children tended to produce a shorter average duration per burst for the unaspirated stops. In addition, these children produced longer VOT-lags than the adults for both unaspirated and aspirated stops. These results indicate that children in this age may not have developed an adult-like airflow control and laryngeal–oral timing pattern in producing stop consonants. The acoustic development of stop sounds is a long-term process that likely extends beyond three years of age. When comparing the current findings on Mandarin-speaking children with previous studies on the acquisition of stop consonants in English-speaking children, the methodological differences as well as the participants' different language backgrounds should be noted. Most of the cited studies collected spontaneous speech samples from very young children and analyzed the VOT data for individual participants. The present study elicited designed speech materials from relatively older children and conducted group analyses. Further, the present study had a relatively small sample size from a limited number of participants. The inconsistent vowel environment between velar stops and the other two places caused by phonotactic constraints may induce a certain confounding effect for the interpretation of statistical results. For future studies, more subjects should be recruited for each age group. In addition, a larger size of spontaneous speech samples should also be collected from each child, which would enable us to examine the distributional characteristics of individual participants in a more natural setting. This information helps us to fully describe the developmental trend of phonetic categories in young children (Hitchcock & Koenig, 2013; Koenig, 2000).

**Acknowledgments.** This research was supported, in part, by the Alumni Grants for Graduate Research and Scholarship at the Ohio State University. I would like to thank the children and their parents for their participation. I would also like to thank Christina Gonzales for her careful reading of this paper, and the anonymous reviewers for their constructive and insightful comments.

## References

- Abramson, A. S. (1977). Laryngeal timing in consonant distinctions. *Phonetica*, 34, 295–303.
- Barton, D., & Macken, M. A. (1980). An instrumental analysis of the voicing contrast in word-initial stops in the speech of four-year-old English-speaking children. *Language and Speech*, 23, 159–69.
- Berenthal, J., Bankson, N., & Flipsen, P. (2009). *Articulation and phonological disorders: speech sound disorders in children*, 6th ed. Englewood Cliffs, NJ: Prentice-Hall.
- Bond, Z. S., & Wilson, H. F. (1980). Acquisition of the voicing contrast by language-delayed and normal-speaking children. *Journal of Speech and Hearing Research*, 23, 152–61.
- Bradshaw, M. (1999). *A crosslinguistic study of consonant-tone interaction* (Doctoral dissertation, Ohio State University). Retrieved from ProQuest Dissertations and Theses (Publication No. AAT 9941291).

- Chao, K. Y., & Chen, L. M.** (2008). A cross-linguistic study of voice onset time in stop consonant productions. *International Journal of Computational Linguistics & Chinese Language Processing*, 13(2), 215–32.
- Chen, L. M., Chao, K. Y., & Peng, J. F.** (2007). VOT productions of word-initial stops in Mandarin and English: a cross-language study. In *Proceedings of the 19th Conference on Computational Linguistics and Speech Processing (ROCLING, 2007)*, retrieved from <http://www.aclweb.org/anthology/O07-2004> (last accessed 29 March 2018).
- Cho, T., & Ladefoged, P.** (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2), 207–29.
- Davis, K.** (1995). Phonetic and phonological contrasts in the acquisition of voicing: voice onset time production in Hindi and English. *Journal of Child Language*, 22(2), 275–305.
- Dodd, B., Holm, A., Hua, Z., & Crosbie, S.** (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, 17(8), 617–43.
- Edwards, J., & Beckman, M. E.** (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics*, 22(12), 937–56.
- Engstrand, O., & Williams, K.** (1996). VOT in stop inventories and in young children's vocalizations: preliminary analyses. *Speech, Music, and Hearing Quarterly Progress and Status Report*, 37(2), 97–9.
- Gandour, J., Petty, S. H., Dardarananda, R., Dechongkit, S., & Mukgoen, S.** (1986). The acquisition of the voicing contrast in Thai: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 13(3), 561–72.
- Gilbert, J. H. V.** (1977). A voice onset time analysis of apical stop production in 3-year-olds. *Journal of Child Language*, 4, 103–13.
- Hitchcock, E. R., & Koenig, L. L.** (2013). The effects of data reduction in determining the schedule of voicing acquisition in young children. *Journal of Speech, Language, and Hearing Research*, 56(2), 441–57.
- Hitchcock, E. R., & Koenig, L. L.** (2015). Longitudinal observations of typical English voicing acquisition in a 2-year-old child: stability of the contrast and considerations for clinical assessment. *Clinical Linguistics & Phonetics*, 29(12), 955–76.
- Hua, Z., & Dodd, B.** (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language*, 27(1), 3–42.
- Imbrie, A. K. K.** (2005). *Acoustical study of the development of stop consonants in children* (Doctoral dissertation, Massachusetts Institute of Technology).
- Keating, P. A.** (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286–319.
- Kewley-Port, D., & Preston, M.** (1974). Early apical stop production: a voice onset time analysis. *Journal of Phonetics*, 2, 195–210.
- Klatt, D. H.** (1975). Voice Onset Time, friction, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686–706.
- Koenig, L. L.** (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research*, 43, 1211–28.
- Koenig, L. L.** (2001). Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends. *Journal of Speech, Language, and Hearing Research*, 44, 1058–68.
- Lee, S. A. S., & Iverson, G. K.** (2008). Development of stop consonants in Korean. *Korean Linguistics*, 14(1), 21–39.
- Lee, S. A. S., & Iverson, G. K.** (2012). Stop consonant productions of Korean–English bilingual children. *Bilingualism: Language and Cognition*, 15(2), 275–87.
- Lee, S. A. S., & Iverson, G. K.** (2017). The emergence of phonetic categories in Korean–English bilingual children. *Journal of Child Language*, 44(6), 1485–515.
- Li, F.** (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *Journal of the Acoustical Society of America*, 133(2), EL142–EL147.
- Lisker, L.** (1986). 'Voicing' in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29, 3–11.
- Lisker, L., & Abramson, A. S.** (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.
- Löfqvist, A.** (1980). Interarticulator programming in stop production. *Journal of Phonetics*, 8, 475–90.



- Löfqvist, A.** (1992). Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants. *Language and Speech*, 35, 15–28.
- Lowenstein, J. H., & Nittrouer, S.** (2008). Patterns of acquisition of native voice onset time in English-learning children. *Journal of the Acoustical Society of America*, 124(2), 1180–91.
- Macken, M. A., & Barton, D.** (1980a). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- Macken, M. A., & Barton, D.** (1980b). The acquisition of the voicing contrast in Spanish: a phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7(3), 433–58.
- Maddieson, I.** (1997). Phonetic universals. In J. Laver & W. J. Hardcastle (Eds.), *The handbook of phonetic sciences* (pp. 619–39). Oxford: Blackwell.
- Morris, R. J., McCrea, C. R., & Herring, K. D.** (2008). Voice onset time differences between adult males and females: isolated syllables. *Journal of Phonetics*, 36(2), 308–17.
- Nittrouer, S.** (1993). The emergence of mature gestural patterns is not uniform: evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research*, 36(5), 959–72.
- Pan, H.** (1995). Longitudinal study of the acquisition of Taiwanese initial stops. *Ohio State University Working Papers in Linguistics*, 45, 131–59.
- Preston, M., & Yeni-Komshian, G.** (1967). Studies on the development of stop-consonants in children. *Haskins Laboratories Status Report on Speech Research*, SR11, 49–52.
- Preston, M. S., Yeni-Komshian, G., Stark, R. E., & Port, D. K.** (1968). Developmental studies of voicing in stops. *Haskins Laboratories Status Report on Speech Research*, SR13/14, 181–4.
- Rochet, B. L., & Fei, Y.** (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105–6.
- Romeo, R., Hazan, V., & Pettinato, M.** (2013). Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America*, 134(5), 3781–92.
- Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A.** (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4), 779–98.
- Smith, B. L.** (1978). Temporal aspects of English speech production: a developmental perspective. *Journal of Phonetics*, 6, 37–67.
- So, L. K. H., & Dodd, B. J.** (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language*, 22(3), 473–95.
- So, L. K. H., & Zhou, J.** (2000). *Putonghua Segmental Phonology Test*. Nanjing: Nanjing Normal University.
- Stevens, K. N.** (2000). *Acoustic phonetics*, Vol. 30. Cambridge, MA: MIT Press.
- Templin, M. C.** (1957). Certain language skills in children: their development and interrelationships. *Institute of Child Welfare Monographs*, 26. Minneapolis, MN: University of Minnesota Press.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S.** (2005). Development of vocal tract length during early childhood: a magnetic resonance imaging study. *Journal of the Acoustical Society of America*, 117(1), 338–50.
- Weismer, G.** (1979). Sensitivity of voice-onset-time (VOT) measures to certain segmental features in speech production. *Journal of Phonetics*, 7, 197–204.
- Wellman, B. L., Case, I. M., Mengert, I. G., & Bradbury, D. E.** (1931). Speech sounds of young children. *University of Iowa Studies: Child Welfare*, 5(2), 82.
- Whiteside, S. P., & Marshall, J.** (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica*, 58(3), 196–210.
- Zlatin, M. A., & Koenigsknecht, R. A.** (1976). Development of the voicing contrast: a comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research*, 19, 93–111.