

# Voice Onset Time in Children and Adults: Evidence from Mandarin Voiceless Stops

Junzhou Ma<sup>1,a</sup>, Xiaoxiang Chen<sup>1,b,\*</sup>, Yezhou Wu<sup>2,c</sup>

<sup>1</sup>School of Foreign Languages, Hunan University, Changsha, Hunan, China

<sup>2</sup>School of Western Languages and Literature, Xi'an International Studies University, Xi'an, Shaanxi China

<sup>a</sup>majunzhou@hnu.edu.cn, <sup>b</sup>laspeech@hnu.edu.cn, <sup>c</sup>wuyezhou0821@163.com

**Keywords:** Voice onset time, children, adult, Mandarin voiceless stops.

**Abstract:** This study seeks to investigate the effect of age and sex on voice onset time(VOT) of stops in Mandarin-speaking children and adults. 13 monolingual children and 13 adults participated in a sentence production experiment. Target sounds were obtained from productions elicited by a sentences reading experiment. Results showed that children had significantly longer VOT values for voiceless aspirated stops in comparison to adults. However, no significant age differences in VOT were observed for voiceless unaspirated stops. In addition, significant sex differences in VOTs were only found for voiceless aspirated stops between adults, whereas no significant sex differences in VOT were found for both aspirated and unaspirated stops produced by children. Possible accounts were discussed.

## 1. Introduction

Voice onset time (VOT) is the time interval between the burst release and the start of the subsequent periodic vibration of vocal folds, and is employed as a primary acoustic cue to differentiate stops consonants in many languages [1].

The vocal fold can vibrate prior to, simultaneous to, or subsequent to the burst release, based on which stop constants fall into three categories, long-lag (60-100ms) stops, short-lag (0-25ms) stops, and pre-voiced (negative values) stops [2] For instance, Mandarin and English show a two-way contrast, short-lag and long-lag stops across three places of articulation: labial, alveolar and velar, while Thai makes a three-way contrast, pre-voiced, short-lag, and long-lag stops. Despite that English and Mandarin both have short-lag and long-lag stops, they do not occupy the same range on VOT continuum. Mandarin stop consonants are phonemically realized as voiceless aspirated stops and voiceless unaspirated stops and aspiration is employed to make a distinction between this cognate pairs. However, English stops differ with their cognate sound via voicing. Mandarin has six stops consonants/p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>, p, t, k/ [3]. and English stop system consists of three aspirated stops and three voiced stops (/ p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>, b, d, g/) [2,4].

Controlling for other factors, stops followed by high vowels are longer than those preceding low vowels [1]. Place of articulation is another factor that has been reported to affect the production of stop consonants. The VOTs values decreases as the tongue moves from the interior to anterior part of the oral cavity [5]. The sequence are velar stops > alveolar stops > bilabial stops. Stops followed by a vowel with higher fundamental frequency will result in longer VOT values than followed by lower fundamental frequency. Since tone is closely related with fundamental frequency, studies have showed that VOT values differ when the following vowel is associated with different tones in tone language [6].

Abundant research has addressed effects of age on VOT. The production of voiceless stops required fine-grained coordination between articulators. Therefore, it poses great challenge for children to master its production. However, with regard to the differences in VOTs between children and adults, no inclusive conclusion has been drawn. Some showed that children aged 4-6 years old has significantly shorter VOTs than adults [7, 8]. While others observed the opposite, especially in children aged 4 years old and younger groups [9, 10].

In addition, the effects of sex on VOTs have been extensively examined. However, most studied sex-related differences between adults, arriving at inconsistent results. Some reported that females have significantly longer VOT values than males [11-15]. While others argued that there existed no significant sex differences in VOT between males and females [5, 16, 17]. With respect to sex differences in children, the literature is relatively scarce. Whiteside and Marshall [14] found that girls aged 11;10 years old had significantly longer VOTs than boys. Whiteside *et al.*, [15] observed longer VOT values for girls compared to boys at the age of 7;10 and 9;10 years old, and 13;2 years old. However, Yu *et al.*, [17] did not reach the same conclusion, suggesting that boys had apparently longer VOTs than girl at the age of 8-10 years old.

Suggested explanations for age and sex differences in VOTs primarily lie in physiological differences between children and adults, and between males and females. Although many studies investigated age- and sex-related differences in VOTs, most were conducted in English. With respect to Mandarin, there are several studies investigating sex differences in VOTs in adults, however, no studies examined VOT differences between children and adults, and sex-related differences in children. Therefore, the purpose of this study was two-fold. First, this study aimed to examine whether age and sex differences in VOTs can be found in Mandarin other than English. If the same results are obtained, it would provide evidence for the universally physiology-based account. If not, age and sex difference in VOTs are possibly caused by other factors, such as speaking styles, cultural factors and other sociophonetic factors. Second, due to the scarcity of studies on Mandarin, this study attempted to bridge this gap and added new evidence to the literature.

Based on previous findings, we formulated the following hypotheses. First, children have longer VOT values compared with adults. Second, significant sex differences in VOTs exist in children and adults respectively.

## 2. Method

### 2.1. Participants

Thirteen children (M = 6, F = 7  $M_{age} = 7.1$ ), and 13 adults (M = 6, F = 7,  $M_{age} = 21.5$ ) participated in a sentence production experiment. All participants were native monolingual speakers of Mandarin. The children were recruited in a primary school and the adults were from a local university. None have been stayed abroad. Participants who spoke dialects were excluded to avoid dialectal effects. None of them reported language, speech or hearing disorders or any other language problems. This was further justified by their parents and teachers.

### 2.2. Stimuli

Voiceless aspirated and unaspirated stops /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>, p, t, k/, each in the initial position of a monosyllable word, were used as target stimuli. To ensure that they occur in the same vowel context and avoid the influence of vowel quality, each stops was followed by the fixed vowel /a/. To eliminate the effect of tonal change, each target word was associated with a fixed High Level tone. In total, six different combinations were formed. Participants were required to read each target stimuli embedded in a carrier phrase, *Qing Ba \_\_\_\_\_ Du Yi Bian* (Please read \_\_\_\_\_ once). Six filler sentences were included to make participants unaware of the purpose and all these sentences were ordered randomly.

### 2.3 Procedures

Participants were first given time to familiarize themselves with the instructions, procedures, and materials. Experimenters made the recordings in a quiet room. Participants sat in front of a portable recorder (SONY ICD-SX1000) with a sampling rate of 44.1 kHz and 16-bit digitization. The recorder was connected to a cardioid condenser microphone (Takstar PCM-5550) placed approximately 20 cm from the mouth of the speaker. Sentences embedding the Target word were written in simple Mandarin on a piece of paper. They were required to read these target sentences

for trial. The experimenters would correct their mispronunciations. In the formal recording, they were asked to read the sentences 4 times naturally, at a normal speed, and not to emphasize any word. In total, 624 target sounds (26 participants  $\times$  6 phrases  $\times$  4 repetitions) were obtained. Sounds that were exaggerated, affected by background noise, or had no clear burst release were discarded. In the end, 583 sounds were coded independently for further analyses.

## 2.4. Statistical analysis

Acoustic analysis of the productions was made on Praat [18]. VOT, in milliseconds, was obtained by measuring the time interval between the burst release and the beginning of the following vowel through waveforms [19]. The start time was marked manually at the first spike of the burst release by a cursor and the end time was marked at the upward-going zero-crossing that introduces the first periodic wave [20]. Additionally, an experienced research assistant analyzed the randomly selected 10% of the productions to guarantee the reliability. The inter-reliability was  $r = 0.99$ ,  $p < 0.001$ , shown by Pearson's product-moment correlation coefficient analysis.

## 2.5. Statistical analysis

Independent  $t$  test or Mann-Whitney  $U$  test were employed to compare the differences in VOT between children and adults, and between sexes in each age group. An alpha level of 0.05 was set as the level of significance. SPSS 22 was used to perform all the statistical analyses.

## 3. Results

Table 1 Mean, Standard Deviation, and  $p$  values between children and adults

	Children	Adults	$p$
	Mean(SD)	Mean(SD)	
/p <sup>h</sup> /	94.02(23.15)	78.37(14.08)	0.000
/t <sup>h</sup> /	99.34(44.09)	81.54(13.97)	0.011
/k <sup>h</sup> /	107.12(26.36)	82.71(16.10)	0.000
/p/	8.47(4.68)	8.55(3.53)	0.521
/t/	9.45(2.77)	9.87(3.30)	0.323
/k/	24.18(9.67)	20.10(4.96)	0.007

Table 1 summarizes the mean, Standard Deviations of VOTs for children and adults, and the  $p$  values as a function of age group and stop. The results of independent  $t$  test or Mann-Whitney  $U$  test showed that there exist significant age differences in VOTs between children and adults for aspirated bilabial, alveolar and velar stops, all with children having significantly longer VOT values compared to adults, and no significant age differences in VOTs were observed for unaspirated stops, except unaspirated velar stop /k/.

Table 2 Mean, Standard Deviation, and  $p$  values between boys and girls

	Boys	Girls	$p$
	Mean(SD)	Mean(SD)	
/p <sup>h</sup> /	92.42( 24.83)	95.56(21.81 )	0.904
/t <sup>h</sup> /	93.13(36.57 )	105.29(50.32 )	0.983
/k <sup>h</sup> /	106.35( 28.51)	107.82( 24.91)	0.650
/p/	9.21( 4.88)	7.62( 4.40)	0.260
/t/	9.91( 2.20)	9.07(3.15)	0.282
/k/	23.89(8.54 )	24.48( 10.88)	0.824

Table 2 displays the means, standard deviation and  $p$  values as a function of gender and stops. Statistical analysis showed that no significant sex differences in VOTs in children were found for both aspirated and unaspirated stops. However, with respect to the sex difference in VOTs in adults, significant sex differences in VOTs were only found for aspirated stops as shown in Table 3

Table 3 Mean, Standard Deviation, and p values between men and women

	Men	Women	<i>p</i>
	Mean(SD)	Mean(SD)	
/p <sup>h</sup> /	73.59(12.22 )	82.26(14.50)	0.021
/t <sup>h</sup> /	76.33(16.02 )	86.00(10.27)	0.011
/k <sup>h</sup> /	75.58(14.74)	88.82(14.86)	0.002
/p/	8.71(1.82)	8.41(4.67 )	0.778
/t/	10.08(2.15 )	9.68(4.07)	0.650
/k/	20.46(5.32)	19.75(4.65 )	0.852

#### 4. Discussion

This study attempted to examine the effects of age and sex differences in VOTs in Mandarin. Results showed that children had significantly longer VOT values compared to adults for voiceless aspirated stops. No significant age differences in VOTs were found for unaspirated stops. Our results are consistent with previous studies in terms of the production of voiceless aspirated long-lag stops [17]. Age differences in voiceless aspirated long-lag stops are possibly due to the relatively more complex articulation and finer temporal coordination process in the production of voiceless aspirated stops over which children have immature control, resulting in their failure to delay glottal vibration [7,10], and children reach adult-like production of voiceless stops much later. Studies have shown that the children's control of motor skills attained adult-like level by 11 years. Compared to voiceless aspirated stops, the articulation of short lag stops is relatively easier and children can produce adult-like short-lag stops before 18 months old [7,8], which in turns explains why no significant age differences in VOTs were found between children and adults in the current study. However, significant age differences for unaspirated velar /k/ was possibly caused by the relative difficulty in producing this sound in comparison to the two anterior unaspirated stops.

With respect to sex effects, there exist no statistically significant differences between boys and girls for both aspirated and unaspirated stops, however, significant sex differences in VOTs were observed between adult males and females for voiceless aspirated stops. Our findings are not consistent with previous research [14,15], which showed that significant sex differences in VOT exist in preadolescent children. A suggested account is no significant physiological differences in vocal tract morphology were found in children in prepuberty. However, with their physical development and maturation, physiological differences between males and female become fairly marked around puberty and continue to adulthood [21] For example, the size of supraglottal vocal tract, vocal tract constriction, and lung volume in females are relatively smaller than males. These factors contribute to sex differences in VOTs in adults. However, the absence of sex differences in VOTs between males and females for unaspirated is unexpected, which is not in consistent with previous studies. A possible account is that the VOT range of unaspirated short-lag stops is 8-20 ms in Chinese children, resulting in smaller acoustic space for variation.

#### 5. Conclusion

The current study provides new evidence for the study of effects of age and sex on VOTs in Mandarin. Although findings confirmed the role of physiological differences, our results were not completely consistent with previous studies in English, indicating that age and sex differences in VOTs are not completely universal, but constrained by other factors such as age, sex and aspiration.

#### Acknowledgments

This project was supported by National Social Science Foundation of China (14BYY144).

## References

- [1] Cho, T. and Ladefoged, P. (1999) Variation and Universals in VOT: Evidence from 18 Languages *Journal of Phonetics*. 27, 207-229.
- [2] Lisker, L. and Abramson, A. S. (1964) A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*. 20, 384-422.
- [3] Rochet, B.L. and Fei Y. (1991) Effect of Consonant and Vowel Context on Mandarin Chinese VOT: Production and Perception. *Canadian Acoustics*. 19, 105–106
- [4] Docherty, G. (1992) *The Timing of Voicing in British English Obstruents*. New York: Foris
- [5] Morris, R.J., McCrea C.R. and Herring, K.D. (2008) Voice Onset Time Differences Between Adult Males and Females: Isolated Syllables. *Journal of Phonetics*. 36,308–317
- [6] Peng, J.F., Chen L.M. and Lee, C.C. (2009). Tonal Effects on Voice Onset Time. *Computational Linguistics and Chinese Language Processing*. 14,341–362.
- [7] Kewley-Port, D. and Preston M.S. (1974) Early Apical Stop Production: A Voice Onset Time Analysis. *Journal of Phonetics*. 2,194–210.
- [8] Macken, M.A. and Barton D. (1980) The Acquisition of the Voicing Contrast in English: Study of Voice Onset Time in Word-Initial Stop Consonants. *Journal of Child Language*. 7, 41–74
- [9] Barton, D. and Macken, M.A. (1980) An Instrumental Analysis of the Voicing Contrast in Word-Initial Stops in The Speech of Four-Year-Old English-Speaking Child. *Language and Speech*. 23,159–169.
- [10] Smith, B.L. (1978) Effects of Place of Articulation and Vowel Environment On Voiced Stop Consonant Production. *Glossa*. 12,163–175
- [11] Koenig, L.L. (2000) Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. *Journal of Speech, Language and Hearing Research* 43,1211–1228.
- [12] Robb, M, Gilbert, H. and Lerman, J. (2005) Influence of Gender and Environmental Setting On Voice Onset Time. *Folia Phoniatica Et Logopaedica*. 57, 125–133.
- [13] Whiteside, S.P. and Irving, C.J. (1998). Speakers' Sex Differences in Voice Onset Time: A Study of Isolated Word Production. *Percept Mot Skills*. 86, 651–654.
- [14] Whiteside, S.P. and Marshall J. (2001) Developmental Trends in Voice Onset Time: Some Evidence for Sex Differences. *Phonetica*. 58,196–210.
- [15] Whiteside, S.P., Henry, L. and Dobbin R. (2004) Sex Differences in Voice Onset Time: A Developmental Study of Phonetic Context Effects in British English. *Journal of The Acoustical Society of America*. 116, 1179–1183.
- [16] Sweeting, P.M. and Baken, R.J. (1982) Voice Onset Time in Normal-Aged Population. *Journal of Speech, Language and Hearing Research*. 25, 129–134.
- [17] Yu, V.Y., De Nil, L.F. and Pang, E.W. (2015) Effects of Age, Sex and Syllable Number On Voice Onset Time: Evidence from Children's Voiceless Aspirated Stops. *Language and Speech*. 58.152–167.
- [18] Boersma, P. and Weenink, D. (2016) PRAAT: Doing Phonetics by Computer, Version Praat 6.0.05.
- [19] Francis, A.L., Ciocca, V. and Yu J.M.C. (2003) Accuracy and Variability of Acoustic Measures of Voicing Onset. *Journal of The Acoustical Society of America*.113,1025–1032.
- [20] Li, F.F. (2013) The Effect of Speakers' Sex On Voice Onset Time in Mandarin Stops. *Journal of The Acoustical Society of America*.133, E1142–E1147.
- [21] Vorperian, H.K., Wang S., Schimek, E.M., et al. (2011). Development of sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, language and Hearing Research*.54, 995–1010.