# A CROSS-LANGUAGE COMPARISON OF VOT DEVELOPMENT IN ENGLISH- AND MANDARIN-SPEAKING CHILDREN AND ADOLESCENTS

*Graham McKenzie and Fangfang Li*

University of Lethbridge, Lethbridge, Alberta, Canada
graham.mckenzie@uleth.ca and fangfang.li@uleth.ca

## ABSTRACT

Voice onset time (VOT) has been shown to vary with speaker's sex, but current cross-language developmental studies lack the ability to explain how sex-specific speech patterns come about independent of a specific-language context. In this research, we examined English- and Mandarin-speaking children and adolescents aged 6 to 17 to determine cross-language similarities and differences. VOT of word-initial /t/ and /d/ was measured and normalized by vowel and word duration to control for speech rate.

We found that in both languages, adolescent boys produced significantly longer VOT for /t/. No age or sex-related difference was found for /d/. The patterns are similar across the two languages. These results suggest that at least part of the sex-specific VOT patterns can be attributed to biological factors, and in our study, most likely anatomical differences that separate the two sexes in adolescence.

**Keywords**: VOT development, cross-language, English, Mandarin

## 1. INTRODUCTION

Human speech simultaneously encodes linguistic information and a variety of other personal dimensions of a speaker, including age, geographic and dialect provenance, educational background and socioeconomic standing [21, 36, 37]. However, by far the most salient variations in adult speech are those that cue a speaker's sex [44].

Sex-specific speech production patterns have been attested in a wide variety of speech sounds in both adult and children's speech. For example, the two sexes are noted for their differentiated ways of producing vowels. In general, women produce vowels with higher voice pitch and formants frequencies than men [12, 36, 40]. These sex-specific vowel production differences are also evident in the speech of children and adolescents [1, 3, 6, 9, 33, 22, 36]. Another example is the sibilant /s/, which has been noted for its sex-differentiated feature [7, 8, 11, 14, 23, 26, 31, 42]. Women articulate /s/ with a more anterior tongue tip constriction in the oral cavity, resulting in more high-frequency components in the fricative noise spectrum, while men's tongue tip positions are further back which yields energy in a lower spectral frequency range. Boys and girls start to diverge in their /s/ production as early as age 4 and 5 years [7, 18, 25]. Boys with gender identity disorders are also found to produce a more anterior variant of /s/ than gender-typical boys [34].

Voice onset time of voiceless stop consonants (/p/ as in *pea*, /t/ as in *tea*, and /k/ as in *key*) are sex-specific [19, 24, 38, 39, 41, 42, 43, 47, 48]. Adult women produce /p/, /t/, and /k/ with longer voice onset time (VOT) than men. VOT is the temporal difference between oral closure and onset of vocal fold vibration and is an effective acoustic parameter in differentiating stop consonants in many languages [5, 27].

By far, most such reports on the sex-specific VOT productions are on English. Although there exist controversies regarding the magnitude of the sex difference, the general consensus is that females produce longer VOT for voiceless aspirated stops (e.g. /p/ as in *pea*, /t/ as in *tea*, and /k/ as in *key*) than males [19, 38, 39, 42, 45, 47, 48].

Interestingly, the sex-related difference in VOT has also been reported in Mandarin Chinese. Similar to English, Mandarin has a two-way contrast system in stop consonants between voiced and voiceless stops that are phonetically realized as voiceless unaspirated and voiceless aspirated stops [4]. Also similar to English, adult females are found to produce longer VOT for voiceless aspirated stops than males [24, 29]. In a recent study that compared the adult speakers of the two languages in their stop productions, the similar trend in sex differentiation between the two languages has been confirmed, along with the finding that the sex difference is more robust in Mandarin than in English [45].

The origin of the sex difference in VOT is not well understood and is likely to be both biological- and socially-conditioned [46, 49]. However, the cross-language similarities found in previous studies suggest more towards the biological nature of this sex-specific speech pattern.

Developmental studies that elucidate the origin of sex-specific VOT patterns in adult speech are scarce in both English and Mandarin. Most studies on VOT development focus on infants or children before school age [13, 15, 17, 30]. Among the studies that

examine older children, the effect of sex is very often neglected [10, 28, 33, 35, 50].

As far as we know, there is only one recent study on the development of English-speaking children and adolescents that is specifically focused on the effect of children's sex [49]. In that study, VOT of word-initial /p/ were examined on 70 English speakers ranged in age from 4 to 18 years old. They found a significant effect of sex on children between age 8 and 11, in which boys are found to produce longer VOT values than girls. For Mandarin, again, only one recent study that specifically investigated the effect of speaker sex on children's VOT productions [29]. Eighty-five children over the age span of 6 to 18 were examined when producing three pairs of Mandarin stops in the labial, alveolar, and velar positions. Different from Yu et al.[49], Ma et al. [29] did not find greater values of VOT in boys' productions. Instead, they reported that females exhibited longer voiceless VOT throughout childhood and adolescents. This effect was less pronounced during the onset of male puberty at age 12 but became stronger into adolescence and adulthood.

The varied results reported in past research could be in part due to a difference in research methodologies, and the small number of participants tested. Often the type of stimulus used in these experiments is inconsistent with each other, making comparison difficult. For example, while Yu et al. [49] and Ma et al. [29] both have similar acoustic stimuli, the speaker's speech rate was not controlled in either experiment. Speech rate has been shown to significantly affect VOT values, with the faster the speaker's speech rate, the shorter the resulting VOT [16]. Without a control on speech rate, it is very difficult to compare these two studies, despite their similarities. Other factors such as vowel-context [39, 49], number of syllables [49], and speaker test-environment [38] have also been shown to affect VOT length. Significant differences in methodologies not only make it difficult to compare within-language developmental variation, but also make it nearly impossible to make cross-language comparisons.

The current study was conducted to systematically compare the effect of speaker sex on the development of VOT across English and Mandarin Chinese. Cross-language developmental studies of this sort with controlled methodology could better reveal similarities shared between languages or differences due to specific language context. It also sheds light into the extent to which common biological factors shape the development in different social context.

The objective of this research is to look at the development of sex differences in VOT in English and Mandarin-Chinese. Both of these languages exhibit sex-specific patterns in VOT in adult speakers [24, 49, 29, 38]. There is also evidence that these patterns emerge in childhood in each language [29, 49]. As mentioned earlier, both Mandarin and English share similar two-way contrast (i.e. short-lag (voiced/voiceless unaspirated) vs long-lag (voiceless aspirated)) contrast, with three similar places of articulation [27]. This contrast makes these two languages ideal for comparison.

This research takes a cross-linguistic approach, with a controlled methodology for age, word-context, and speaker's speech rate. These controls allow for comparison within language through development from childhood into late-adolescence, and also across both languages. This research will add to the body of literature about the role of sex/gender on the development of speech from a cross-language perspective. Importantly, the cross-linguistic approach taken helps contribute to our understanding of the possible biological and social mechanisms of VOT development, and their interaction with a child's sex and gender.

## 2. METHODS

### 2.1. Participants and task

A total of 137 English-speaking and 260 Mandarin-speaking children and adolescents participated in the study. They were divided into three age groups: ages 6~8, 11~13, 15~17 years for both English and Mandarin groups. English-speaking participants were recruited and tested in Lethbridge, Alberta, Canada, while Mandarin-speaking participants were tested in Luoyang, Henan, China.

The breakdown of participants based on their age and sex can be found in Table 1. Participants were engaged in a word-elicitation task, during which they were seated in front of a computer screen and were asked to produce the words beginning with /t/ or /d/ in various vocalic contexts when prompted by pictures describing those words. Participant's voices were recorded using Marantz 661 digital voice recorder and a Shure SM58 unidirectional microphone. The recordings were made using a 44.1 kHz sampling rate and at a 16-bit quantization.

For each language, we chose words that are familiar to children. Specifically, a total of 18 words were selected for English and Mandarin, for the two target consonants, in front of three vowel contexts (i.e., /i/, /ae/, and /u/). The example words are "toothpaste" vs. "doodle" in English, and /tu.zi/ ("rabbit") vs. /du.zi/ ("belly") in Mandarin Chinese.

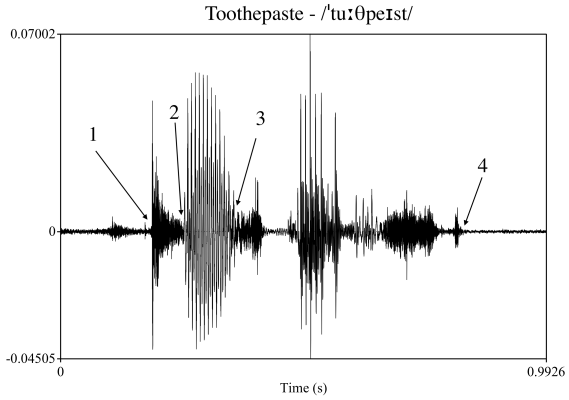
**Table 1**: A breakdown of participants by language, age group, and sex.

| | ENGLISH | | MANDARIN | |
|---|---|---|---|---|
| AGE | *Female* | *Male* | *Female* | *Male* |
| 6-8 | 39 | 36 | 54 | 54 |
| 11-13 | 14 | 13 | 39 | 38 |
| 15-17 | 16 | 12 | 39 | 33 |

### 2.3. Acoustic analysis

The acoustic analysis was performed using PRAAT software [2], following the established procedure in published studies [20]. Specifically, four temporal points of interest were measured: burst onset, vowel onset, vowel end, and word end (see figure 1). Vowel end and word end were used to calculate vowel duration and word duration respectively. These measures were used to control for speech rate.

**Figure 1**: The four measurement points for voice onset time for the word "toothpaste". Points of interest recorded: burst onset (1), vowel onset (2), vowel end (3), and word end (4).




### 2.4. Statistical analysis

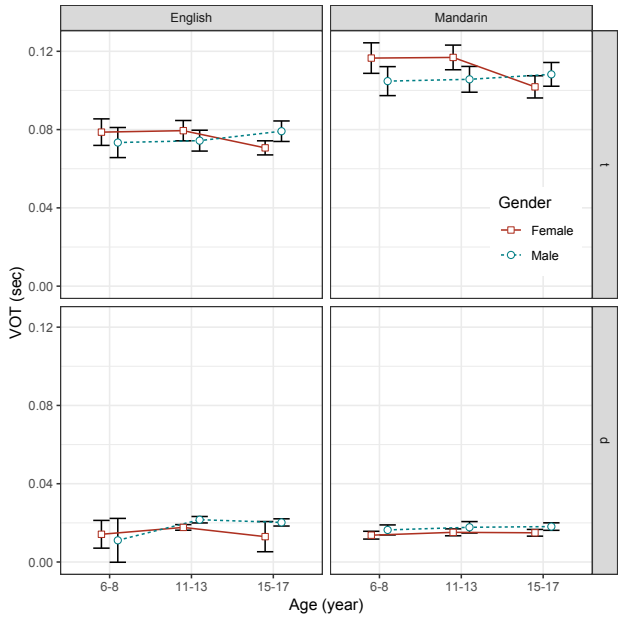
Two linear mixed effects (LME) models were conducted, one for each language, with the dependent variable being raw VOT values. The independent fixed effects variables were target consonant (/t /vs. /d/), age group (6-8, 11-13, and 15-17), speaker's sex (male vs female), and possible interaction terms among the three variables, as well as vowel duration, and word duration to control for speech rate. Individual variation was taken care of by including a random effect of speakers.

## 3. RESULTS

The results of the LME model for English VOT revealed a main effect of sex (t=-2.86, p=0.004), and a significant interaction between sex and the oldest age group (t=2.20, p=0.027), which suggests that adolescent boys and girls have different VOT patterns. The results for Mandarin revealed a significant interaction between sex and the target consonant /t/ (t=-7.56, p<0.001), suggesting that the way that boys and girls differ in their VOT is the not the same across the two target consonants. Furthermore, we found a significant interaction between sex, target /t/, and the 15-17 age group (t=2.20, p=0.027).

Figure 2 illustrates this interaction term graphically for the two languages. From the figure, it is clear that for English speakers, girls tend to longer VOT for both targets in the youngest age group and this sex-related pattern is reversed in the oldest age group. For Mandarin speakers, a similar switch-over in the VOT patterns in the adolescent group was observed, but only in the target /t/ sound. For target /d/, boys consistently produce longer VOT than girls for the target /d/ for all three age groups.

**Figure 2**: Age and sex-related changes in voice onset time (normalized by vowel duration) for English and Mandarin children and adolescents.




## 4. DISCUSSION

The current study systematically compared English and Mandarin Chinese for the effect of sex on VOT in children' and adolescent's stop productions. Such a cross-language study could avoid confounds due to methodological inconsistencies and thus help reveal cross-language similarities or differences that could light on the origin of sex-specific VOT pattern.

Our results indicate a consistent developmental trend for both languages with respect to the effect of sex on both stops. For English, adolescent boys were

found to produce longer VOT values than girls. This result is similar to what Yu et al. [49] reported that boys' stops are more aspirated than girls. However, Yu et al. [49] found that significant sex difference only exists in 9-year-olds, while ours found the difference in the 15~17 age group. Despite the specific age range that shows sex difference differs between Yu et al. [49] and our study, the fact that we both found boys' VOT are longer than girls suggests that there is a switch-over in the pattern of sex difference in adulthood, as adult males were reported to produce shorter VOT than females for both languages [24, 29].

Although developmental pattern in Mandarin-speaking children and adolescents does not exactly mirror that was found in English speakers, in particular for voiced stops, a similar switch-over in adolescents was observed for voiceless stops. This cross-language similarity leads us to propose that the origin of such difference is more biological-based. We speculate that the more aspirated stops produced by adolescent boys in both languages for /t/ may be due to the changing larynx that boys go through during puberty. However, further research needs to be done to ascertain the developmental pattern and underlying mechanisms.

It is worth noting some cross-language differences as well. One such difference is that the effect of sex in adolescents' /t/ production is more robust in Mandarin Chinese than English, as shown in Figure 2. This finding aligns well with investigations on sex difference in VOT in adults in the two languages, where Mandarin speakers were found to produce a more sex-differentiated /t/ than English speakers [45].

There are still more puzzle pieces to be added to the overall picture before definite conclusions can be drawn about the origin and mechanisms of sex difference in stop consonants. For example, it is unclear why Yu et al. [49] and our study found different switch-over point for English-speaking children, and why the pattern eventually reversed in adulthood for both languages. Further work that brings in more information and perspectives with respect to how biology interacts with social factors will hopefully answer these questions.

## 5. REFERENCES


[1] Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231-238.

[2] Boersma, P. & Weenink, D. (2017). PRAAT. Institute of Phonetic Sciences University of Amsterdam, The Netherlands. Free software retrieved in 2017 from http://www.fon.hum.uva.nl/praat/.

[3] Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*, *97*(4), 2603-2606.

[4] Chao, K. Y., & Chen, L. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing,* 13(2), 215-232.

[5] Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 17 endangered languages. *Journal of Phonetics*, *27*, 207-229.

[6] Eguchi, S. & Hirsh I.J. (1969) Development of speech sounds in children. Acta Otolaryngologica Supplementum. 257:307–356.

[7] Fox, R., & Nissen, S. (2005). Sex-Related Acoustic Changes in Voiceless English Fricatives. Journal Of Speech Language And Hearing Research, 48(4), 753

[8] Fuchs, S., & Toda, M. (2009). Do differences in male versus female /s/ reflect biological factors or sociophonetic ones? In S. Fuchs, Toda, M. & Zygis, M. (Ed.), *An interdisciplinary guide to turbulent sounds*. Berlin: Mouton de Gruyter.

[9] Hasek, C. S., Singh, S., & Murry, T. (1980). Acoustic attributes of preadolescent voices. *The Journal of the Acoustical Society of America*, *68*(5), 1262-1265.

[10] Hazan, V., Romeo, R., & Pettinato, M. (2013). The impact of variation in phoneme category structure on consonant intelligibility. Proceedings of Meetings on Acoustics, 060103–060103.

[11] Heffernan, K. (2004). Evidence from HNR that /s/ is a social marker of gender. Toronto working papers in Linguistics 23(2). 71–84.

[12] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099-3111.

[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), 441-457.

[14] Johnson, K, (1991) Differential effects of speaker and vowel variability on fricative perception. Language and Speech, 34(3), 265–279.

[15] Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19*(3), 421-447.

[16] Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2), 117-128.

[17] Kewley-Port, D. & Preston, M. S. (1974) Early apical stop production: A voice onset time analysis. Journal of Phonetics, 2,195–210.

[18] Kinsman, M. & Li, F. (2013) The relationship between gender-differentiated productions of /s/ and gender role behaviour in young children. *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech)*, Lyon, France, pp.1283-1286.

[19] 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*(5), 1211-1228.

[20] Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. Journal of Phonetics, 40(6), 725-744.

[21] Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

[22] Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455-1468.

[23] Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese* (Doctoral dissertation, The Ohio State University).

[24] Li, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America*, *133*(2), EL142–EL147.

[25] Li, F., Rendall, D., Vasey, P., Kinsman, M., Ward-Sutherland, A., & Diano, G. (2016). The development of sex/gender-specific /s/ and its relationship to gender identity in children and adolescents. *Journal Of Phonetics*, 57, 59-70.

[26] Linville, S. (1998) Acoustic correlates of perceived versus actual sexual orientation in men's speech. Pholia Phoniatrica et Logopaedica 50, 35–48.

[27] Lisker, L., & Abramson, A.S. (1964) A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20*, 384-422.

[28] Lowenstein, J. H., & Nittrouer, S. (2008). Patterns of acquisition of native voice onset time in English-learning children. The Journal of the Acoustical Society of America, 124(2), 1180–1191.

[29] Ma, J., Chen, X., Wu, Y., & Zhang, L. (2017). Effects of age and sex on voice onset time: Evidence from Mandarin voiceless stops. *Logopedics Phoniatrics Vocology*, *19*, 1-7.

[30] Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*(1), 41-74.

[31] Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, 34(2), 202-240.

[32] Munson, B., Crocker, L., Pierrehumbert, J., Owen-Anderson, A., & Zucker, K. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. The Journal Of The Acoustical Society Of America, 137(4), 1995-2003

[33] Murry, T., & Singh, S. (1980). Multidimensional analysis of male and female voices. The Journal of the Acoustical Society of America, 68(5), 1294-1300.

[34] 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), 959-972.

[35] Perry, T., Ohde, R., & Ashmead, D. (2001). The acoustic bases for gender identification from children's voices. The Journal Of The Acoustical Society Of America, 109(6), 2988-2998.

[36] Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175-185.

[37] Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*, 10-13.

[38] Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica Et Logopaedica*, *57*(3), 125-133.

[39] Ryalls, J. Zipprer, A. & Baldauff, P. A. (1997) Preliminary investigation of the effects of gender and race on voice onset time. Journal of Speech, Language, and Hearing Research, 40, 642-645.

[40] Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. Language Attitudes: Current Trends And Prospects, 74-84.

[41] Scharf, G.& Masure, H. (2002) Voice onset time in normal speakers of a German dialect: Effects of age, gender and verbal material," in Investigations in Clinical Phonetics and Linguistics, edited by F.Windsor, M. L. Kelly, and N. Hewlett (Erlbaum Associates, Mahwah, NJ), pp. 327–339.

[42] Stuart-Smith, J. (2007). Empirical evidence for gendered speech production:/s/in Glaswegian.

[43] Swartz, B. L. (1992) Gender difference in voice onset time. *Perceptual and Motor Skills, 75,* 983-992.

[44] Trudgill, P. (1983). *On dialect. Social and geographical perspectives*. Oxford: Blackwell.

[45] Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147-161.

[46] Whiteside, S. P., & Irving, C. J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. *Perceptual and motor skills*, *85*(2), 459-463E.

[47] Whiteside, S. P., & Irving, C. J. (1998). Speakers' sex differences in voice onset time: a study of isolated word production. Perceptual and motor skills, 86(2), 651-654.

[48] Whiteside, S. P., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *The Journal of the Acoustical Society of America*, *116*(2), 1179-1183.

[49] Yu, V., De Nil, L., & Pang, E. (2014). Effects of Age, Sex and Syllable Number on Voice Onset Time: Evidence from Children's Voiceless Aspirated Stops. *Language And Speech*, 58(2), 152-167.

[50] 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, Language, and Hearing Research, 19*(1), 93-111.