The effect of speakers' sex on voice onset time in Mandarin stops

Fangfang Li
Department of Psychology, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1J 3M4, Canada
fangfang.li@uleth.ca

Abstract: The goal of the present study is to examine the effect of speakers' gender on voice onset time in Mandarin speakers' stop productions. Word-initial lingual stops were elicited from 10 male and 10 female Mandarin speakers using a word-repetition task. The results revealed differentiated voice onset time (VOT) patterns between the two genders for all four lingual stops on raw VOT values. After factoring out speech rate variation, gender-related differences remained for voiced stops only with females' VOTs being shorter than males. The results, together with previous findings from other languages, suggest a sociolinguistic/stylistic account on the relation between gender and VOT that vary in a language-specific manner.

PACS numbers: 43.70.Bk, 43.70.Kv [AL]
Date Received: September 26, 2012 **Date Accepted:** January 2, 2013

1. Introduction

Voice onset time (VOT), the temporal difference between a stop closure release and the onset of vocal fold vibration, is an effective acoustic parameter in differentiating stop consonants in many languages (Lisker and Abramson, 1964; Cho and Ladefoged, 1999; among others). Along the VOT dimension, stop categories can be roughly classified as having lead (−120–0 ms), short-lag (0–25 ms), and long-lag (40–100 ms) VOTs, although the phonetic implementations of the voicing categories vary with languages. For example, both English and French make a two-way distinction between voiced and voiceless stops, but the English voiceless stop is phonetically realized as long-lag VOT while the French voiceless stop falls within the short-lag VOT range.

VOT is also known to vary with a number of factors, including phonetic context, place of articulation, speech tempo, and speakers' gender. All else being equal, stops have longer VOTs when followed by high vowels as compared to low vowels (Morris *et al.*, 2008; Whiteside *et al.*, 2004). VOT values are also contingent on place of articulation such that the further posterior the constriction, the longer the VOT. Thus velar stops show longer VOTs than bilabial stops (Cho and Ladefoged, 1999; Morris *et al.*, 2008). In addition, VOT values of voiceless stops but not voiced stops are negatively correlated with speech rate (Kessinger and Blumstein, 1997; Pind, 1995).

In English, among the various factors conditioning VOT variations, speakers' gender is the one that has been most frequently attested. For voiceless stops, most studies agree that females display longer VOTs than males (Koenig, 2000; Robb *et al.*, 2005; Ryalls *et al.*, 1997; Swartz, 1992; Whiteside *et al.*, 2004; Whiteside and Irving, 1998). For voiced stops, however, the results are rather mixed. The majority of studies have demonstrated that females produce longer VOT values than males (Koenig, 2000; Robb *et al.*, 2005; Ryall *et al.*, 1997; Swartz, 1992), while Whiteside and Irving (1998) found the opposite. Further, previous studies disagree as to whether the reported sex-related differences for the voiceless stops could be accounted for by varied speech rate between the two sexes. After controlling for speech tempo differences, Swartz (1992) still found significant longer VOT values for females, while Morris, McCrea, and Herring (2008) did not.

Beyond these controversies, a number of interpretations have been offered for the sex-VOT relation in English. For example, Koenig (2000) proposed an account based on anatomical and physiological differences, in particular the trans-glottal air flow difference between men and women. Others have speculated that linguistic, stylistic, and sociolinguistic factors could also play a role (Swartz, 1992; Whiteside and Irving, 1998). If sex-differentiated VOT patterns are truly the result of "biological" processes, other languages sharing similar sound types are expected to exhibit similar tendencies. If, however, such patterns are found to be unique to English, factors specific to a language, such as sociolinguistic or stylistic issues, are more likely to be at the root. Unfortunately, few studies have compared male and female VOT realizations in languages other than English. One study that focuses on speakers' gender and VOT in a different language is Oh (2011), who studied Korean voiceless stops, which contrast three phonation types: fortis, lenis, and aspirated. She found a completely different pattern from that in English: Females produce shorter VOTs than males for the long-lag aspirated stops. This sex-related difference has been interpreted to reflect an ongoing sound change led by women in Korean that merges lenis and aspirated stops. Scharf and Masure (2002) is another study about VOT and gender, but this time in German. Similar to English, German makes a binary distinction between voiceless and voiced stops with voiced stops having short-lag VOTs and voiceless ones having long-lag VOTs. The results of that study indicated a greater distance between the two voicing categories in women's speech. However, no comparisons between men and women were made for individual stop categories.

The purpose of the present study is to examine the relationship between sex and VOT in another language that has similar phonological contrasts to English: Mandarin Chinese. As with English and German, Mandarin stops also make a two-way distinction between short-lag and long-lag VOTs across all three places of articulation: Labial, alveolar, and velar. Further, similar VOT interactions with place of articulation and phonetic context have been found in Mandarin Chinese. That is, Mandarin has longer VOTs for stops when followed by high vowels than by low vowels, and longer VOTs for velar stops than for alveolars (Rochet and Fei, 1991; Chao and Chen, 2008). In contrast to English and German, however, to my knowledge there are no studies in Mandarin that reported the effect of speakers' gender. Thus the question remains as to whether an effect of speakers' gender similar to that in English can be found in Mandarin Chinese or not. The present research addresses this question with the goal of describing the nature of the sex-VOT relationship from a cross-language perspective.

2. Methods

2.1 Participants and task

Twenty adults (10 females and 10 males) aged 18–30 yr were recruited and tested in Songyuan, China. None of them exhibited any previous speech, hearing, or language problems or spoke languages other than standard Mandarin Chinese. During testing, they were seated in front of an IBM laptop computer and were instructed to repeat each word after the audio prompts played out loud through speakers attached to the sides of the computer. At the same time, picture prompts of target words were presented simultaneously with the audio prompts. Participants' speech production was recorded with a Shure dynamic unidirectional microphone into Marantz 660 digital recorder. The sampling frequency was 44.1 kHz with 16 bit digitization.

2.2 Materials

Words beginning with lingual stop consonants (/t/, /d/, /k/, and /g/) followed by the vowels /a/, /u/, and /i/ were selected. No words with an initial CV combination of /gi/ or /ki/ were included because the two sequences are phontactically illegal in Mandarin Chinese. All words were disyllabic, the most common word type in contemporary

Mandarin Chinese. Two word tokens were elicited for each consonant-vowel (CV) combination, yielding a total of 20 words per subject.

2.3 Procedure

Speakers' productions were segmented and speech events such as stop bursts and voicing onsets were labeled using the speech software PRAAT (Boersma and Weenink, 2005). Speech recordings were displayed with a two-tier window, a top tier for waveforms, and a bottom tier for spectrograms. The spectrogram display was set with a view range between 0 and 5000 Hz, 0.005 second window length, and 70 dB dynamic range. In making measurements, burst onset was marked at the peak of the first spike of a cluster of transient noise that constitutes the stop burst. The voicing onset was marked as the first deviation of periodic glottal pulse. The end of word was determined by both the decease of glottal pulsing of the word-final vowels in the waveform tier and accompanying fading of formants in the spectrogram tier. Word duration was defined as the time difference between burst onset and word end. Voice onset time was calculated by taking the time difference between a stop burst onset and the following voicing onset. Ten percent of the original data were re-marked by a different individual as well as by the same individual 6 mon after doing the original measurement. The intra- and inter-rater reliability was 0.96 and 0.98, respectively, as assessed by Peason's correlation coefficient.

3. Results

3.1 Effect of speakers' sex

Table 1 displays the mean VOTs for each consonant in varying vocalic context produced by males and females. The means and standard deviations of VOT are also graphically plotted in Fig. 1 for both sexes and in different vowel contexts. It is clear from both Table 1 and Fig. 1 that for the voiced stops /d/ and /g/, females produced smaller VOT values than males. For the voiceless stops /t/ and /k/, the opposite pattern was found with females' VOTs being longer than males'. Further, for /d/, the VOT values in the context of vowel /a/ were higher than those for the other two vowels. For the velar stops, the VOTs for vowel /a/ and vowel /u/ were different for both sexes.

To determine whether the observed patterns are statistically significant, four separate repeated measure ANOVAs were fitted for each stop consonant. For each individual ANOVA, the dependent variable is VOT, and the independent variables are speakers' sex (between-subject) and vowel (within-subject). An alpha level of 0.05 was set as the level of significance. The results of the four ANOVAs are displayed in Table 2. It is evident from Table 2 that a main effect of speakers' gender was found in all four models, suggesting differentiated VOT values between males and

Table 1. Mean VOT values (in milliseconds) for each lingual stop consonant in varying vowel contexts separated by gender.

		Vocalic context			
Stop consonant	Speakers' sex	/a/	/u/	/i/	Mean
/d/	Female	10.9	12.0	14.7	12.5
	Male	15.8	16.7	20.1	17.5
/t/	Female	94.2	90.1	95.6	93.3
	Male	72.8	79.4	81.3	77.9
/g/	Female	18.4	26.5	NA	22.5
	Male	25.1	34.0	NA	29.5
/k/	Female	89.2	92.4	NA	90.8
	Male	69.3	88.9	NA	78.8

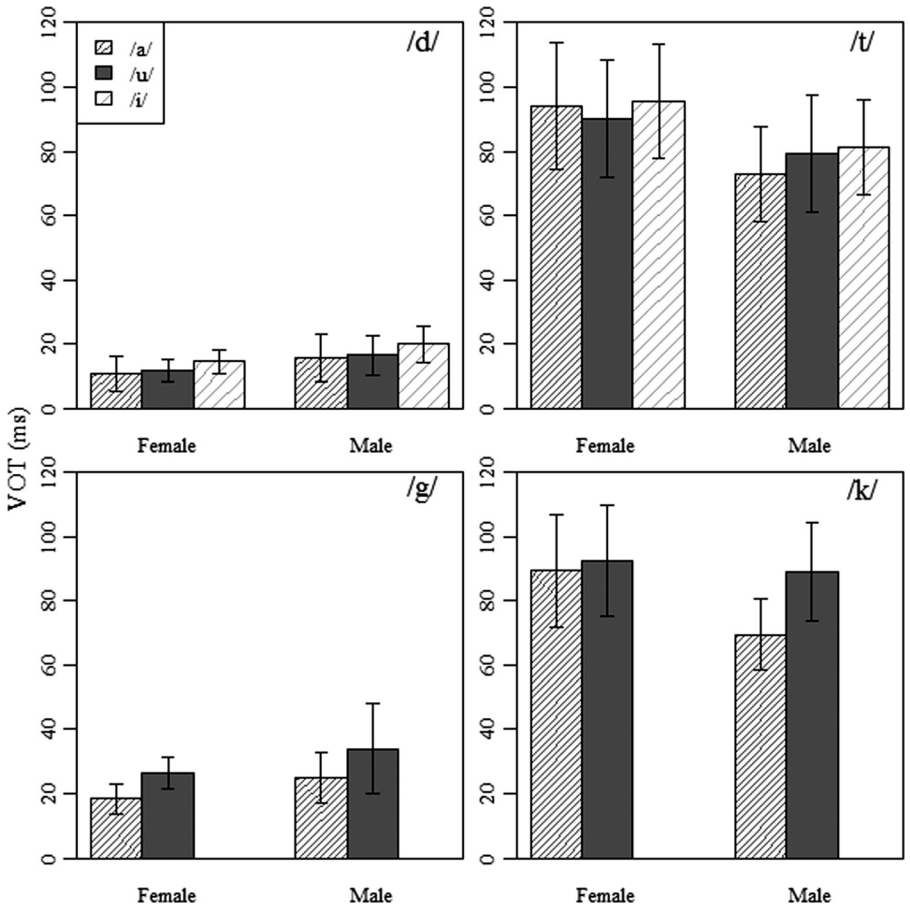

Fig. 1. Bar plots of means for each lingual stop as a function of speakers' sex and vocalic context. Error bar is 1 s.d.

females for these four stop consonants. Second, a main effect of vowel context was found for /d/, /g/, and /k/, but not for /t/. Tukey's HSD *post hoc* test was used to determine the nature of vowel effect for /d/ as there were more than two vowels involved. The *post hoc* test revealed that the VOT values in the vowel /i/ context were

Table 2. Results for individual repeated measure ANOVAs conducted for each stop consonant.

Stop consonant	Effects	F value	P value	Effect size
/d/	Sex	$F(1,17) = 11.79$	$p = 0.003$	$\eta^2 = 0.18$
	Vowel	$F(2,36) = 11.79$	$p < 0.001$	$\eta^2 = 0.9$
	Sex*vowel		n.s.	$\eta^2 = 0$
/t/	Sex	$F(1,17) = 9.04$	$p < 0.001$	$\eta^2 = 0.17$
	Vowel		n.s.	$\eta^2 = 0.02$
	Sex*vowel		n.s.	$\eta^2 = 0.02$
/g/	Sex	$F(1,17) = 7.17$	$p = 0.015$	$\eta^2 = 0.15$
	Vowel	$F(1,18) = 55.2$	$p < 0.001$	$\eta^2 = 0.20$
	Sex*vowel		n.s.	$\eta^2 = 0$
/k/	Sex	$F(1,17) = 5.55$	$p = 0.031$	$\eta^2 = 0.13$
	Vowel	$F(1,18) = 10.45$	$p < 0.001$	$\eta^2 = 0.14$
	Sex*vowel	$F(1,18) = 6.88$	$p = 0.017$	$\eta^2 = 0.07$

significantly larger than those in the vowel /a/ context ($p < 0.001$). The interaction effect between vowel and speakers' gender for /k/ is depicted in Fig. 1 where males show more varied VOTs between the two vowel contexts than females.

3.2 Speaking rate and speakers' sex

Four stepwise multiple linear regressions were carried out to determine whether the sex-related differences were artifacts of differences in speaking rate between men and women for each target consonant. Each regression model has VOT as the dependent variable and includes two independent variables: speakers' sex and averaged word duration for each subject. Results for models of the voiceless stops, /t/ and /k/, only revealed a significant positive correlation between word duration and VOT. That is, the longer a word, the longer the VOT. More importantly, speakers' sex did not account for any significant portion of variation in VOT when speaking rate was controlled for. For the voiced velar stop /g/, word duration did not explain any significant amount of variations in VOT. Also, speaker's sex only did so marginally ($p = 0.07$). For the voiced alveolar stop /d/, speakers' sex instead of word duration, correlated significantly with VOT values ($p = 0.005$, $r^2 = 36\%$).

4. Discussion

In summary, significant gender-related differences were found for all four lingual stops when raw VOT values were examined. Specifically, females produced longer VOTs than males for voiceless stops and shorter VOTs for voiced stops. However, males did talk faster than females for voiceless stops as assessed by word duration. When speech rate was controlled for, a significant effect of speakers' gender was found only for the voiced category. The results are not in agreement with the major pattern reported in English—that females produce longer VOTs than males for both voiceless and voiced stops.

These patterns appear to be best explained by sociolinguistic factors, which tend to be language-specific for three reasons. First, the results from previous studies in relation between sex and voiced stops in English are mixed, among which the majority report either no sex-related differences or the opposite pattern to that in Mandarin, with females' VOTs being longer than males. Crucially, Kong and Weismer (2010) reported longer VOTs in the production of voiced stops for males than females in an English study that used exactly the same procedure as the present one (word-repetition task, 20 subjects, and lingual stops) as both datasets were collected under the same protocol for a cross-language study on phonological acquisition (Edwards and Beckman, 2008). We can conclude, then, that the different patterns for sex-VOT interactions in English and Mandarin Chinese are likely not due to procedural discrepancies. Second, the results of the present study on Mandarin, of other previous studies on English, and the study of Oh (2011) on Korean, are all different. Oh (2011) demonstrates that after factoring out speech rate variation, females produce shorter VOTs for voiceless aspirated stops, a pattern in sharp contrast to both English and Mandarin Chinese. Third, by producing shorter VOT values for voiced stops, the phonemic contrast in voicing is enhanced. The shorter VOT exhibited in females is therefore consistent with the sociolinguistic findings that females tend to adopt a more careful speech style than males (Byrd, 1994; Whiteside, 1996).

To conclude, the present study provides novel evidence demonstrating the relationship between speakers' sex and VOT in Mandarin Chinese. Given that Mandarin shares similar phonological contrast in stops with English, findings in Mandarin can shed light on the nature of sex-VOT relation that has been extensively reported in English. The fact that speakers' sex affects VOT differently in the two languages suggests that language/cultural factors (i.e., sociolinguistic, stylistic effects) rather than biological/anatomical factors may be the cause of why women and men produce stops differently. Future studies with larger sample size and more participants are needed to validate the results reported in the present article on Mandarin Chinese. Further, more

investigations on other languages are needed to allow for a complete characterization of the interplay between VOT and speakers' sex.

Acknowledgments

The work was reported by NSF Grant BCS0739206 to Mary Beckman, by an Ohio State University Department of Linguistics Targeted Investment Award to F.L. and Eun Jong Kong, by NIDCD Grant 02932 to Jan Edwards, and by University of Lethbridge Start-up Fund to F.L.

References and links

Boersma, P., and Weenink, D. (**2005**). "PRAAT: Doing phonetics by computer," version praat 4.3.07.
Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **15**, 39–54.
Chao, K.-Y., and Chen, L.-M. (**2008**). "A cross-linguistic study of voice onset time in stop consonant productions," Comput. Linguist. Chin. Lang. Process. **13**(2), 215–232.
Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phonet. **27**, 207–229.
Edwards, J., and Beckman, M. E. (**2008**). "Methodological questions in studying phonological acquisition," Clin. Linguist. Phonet. **22**(12), 939–958.
Kessinger, R. H., and Blumstein, S. E. (**1997**). "Effects of speaking rate on voice-onset time in Thai, French, and English," J. Phonet. **25**, 143–168.
Koenig, L. L. (**2000**). "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds," J. Speech, Lang. Hear. Res. **43**, 1211–1228.
Kong, E., and Weismer, G. (**2010**). "Correlation of acoustic cues in stop productions of Korean and English adults and children," J. Korean Soc. Speech Sci. **2**(4), 29–37.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustic measurements," Word **20**, 384–422.
Morris, R. J., McCrea, C. R., and Herring, K. D. (**2008**). "Voice onset time differences between adult males and females: Isolated syllables," J. Phonet. **36**, 308–317.
Oh, E. (**2011**). "Effects of speaker gender on voice onset time in Korean stops," J. Phonet. **39**, 59–67.
Pind, J. (**1995**). "Speaking rate, voice-onset time, and quantity: The search for higher-order variants for two Icelandic speech cues," Percept. Psychophys. **57**, 291–304.
Robb, M., Gilbert, H., and Lerman, J. (**2005**). "Influence of gender and environmental setting on voice onset time," Folia Phoniatr. Logop. **57**, 125–133.
Rochet, B. L., and Fei, Y. (**1991**). "Effect of consonant and vowel context on Mandarin Chinese VOT: Production and perception," Can. Acoust. **19**(4), 105–106.
Ryalls, J., Zipprer, A., and Baldauff, P. A. (**1997**). "Preliminary investigation of the effects of gender and race on voice onset time," J. Speech, Lang. Hear. Res. **40**, 642–645.
Scharf, G., and Masur, 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.
Swartz, B. L. (**1992**). "Gender difference in voice onset time," Percept. Mot. Skills **75**, 983–992.
Whiteside, S. P. (**1996**). "Temporal-based acoustic phonetic patterns in read speech: Some evidence for speaker sex differences," J. Int. Phonet. Assoc. **26**, 23–40.
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," J. Acoust. Soc. Am. **116**(2), 1179–1183.
Whiteside, S. P., and Irving, C. J. (**1998**). "Speakers' sex differences in voice onset time: A study of isolated word productions," Percept. Mot. Skills **86**, 651–654 (1998).